

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model.

Run on: April 29, 2003, 17:10:24 ; Search time 11.7002 Seconds

(without alignments)
3367.676 Million cell updates/sec

Title: US-10-009-332-1

Perfect score: 5164

Sequence: 1 MLLGIITLAFAGTAGGFE.....DQCNLHRKPQELDFCVLRPC 950

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2482.5	48.1	968	1 ATSL_MOUSE	P97857 mus musculus
2	2480.5	48.0	967	1 ATSL_HUMAN	Q9uh18 homo sapien
3	2471.1	47.9	967	1 ATSL_RAT	Q9wug1 rattus norv
4	2135.5	41.4	890	1 ATSL_HUMAN	Q9up79 homo sapien
5	2113.4	40.9	905	1 ATSL_MOUSE	P57110 mus musculus
6	1965.8	38.1	1629	1 ATSL_HUMAN	Q9p2h4 homo sapien
7	1917.7	37.1	837	1 ATSL_HUMAN	Q75173 homo sapien
8	1856.3	35.9	930	1 ATSL_MOUSE	Q9r001 mus musculus
9	1847.5	35.8	930	1 ATSL_HUMAN	Q9una0 homo sapien
10	1745.5	33.8	630	1 ATSL_RAT	Q9esp7 rattus norv
11	1396.2	25.1	1077	1 ATSL_HUMAN	Q9h324 homo sapien
12	1275.7	24.7	1593	1 ATSL_HUMAN	P58397 homo sapien
13	1230.3	23.8	1211	1 ATSL_HUMAN	P78325 homo sapien
14	1212.5	23.5	997	1 ATSL_HUMAN	Q95450 h adamts-2
15	1192.2	23.1	1205	1 ATSL_BOVIN	Q9ukp4 homo sapien
16	1183.2	22.9	1205	1 ATSL_HUMAN	P79331 b adamts-2
17	1048.8	20.3	860	1 ATSL_HUMAN	O15072 homo sapien
18	601.5	11.6	245	1 ATSL_BOVIN	Q9ukp5 homo sapien
19	599.9	11.6	207	1 ATSL_BOVIN	Q9tt93 bos taurus
20	381.5	7.4	450	1 ATSL_MOUSE	Q9tt92 bos taurus
21	380.5	7.4	824	1 AD08_HUMAN	P58459 mus musculus
22	362.7	7.0	776	1 AD28_MACFA	P78325 homo sapien
23	362.7	7.0	956	1 AD19_HUMAN	Q9xsl6 macaca fasc
24	349.5	6.8	813	1 AD33_HUMAN	Q9h013 homo sapien
25	344.5	6.7	571	1 DLST_BOVTA	Q9bz11 homo sapien
26	342.5	6.6	775	1 AD28_MOUSE	P30431 bothrops ja
27	340.6	6.6	774	1 AD28_HUMAN	Q9jln6 mus musculus
28	337.7	6.5	857	1 AD22_MOUSE	Q9ukq2 homo sapien
29	332.6	6.4	906	1 AD22_HUMAN	Q9r1v6 mus musculus
30	329.5	6.4	920	1 AD19_MOUSE	Q9p0k1 homo sapien
31	329.5	6.4	903	1 AD12_MOUSE	Q35674 mus musculus
32	321.6	6.2	826	1 AD08_MOUSE	O61824 mus musculus
33	314.6	6.1	909	1 AD12_HUMAN	Q05910 mus musculus
					O43184 homo sapien

RESULT 1

ID	ATSL_MOUSE	STANDARD;	PRT;	968 AA.
AC	P97857; O54768;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase			
DE	with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).			
GN	ADAMTS1			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/SvJ;			
RX	MEDLINE=98110583; PubMed=9441751;			
RA	Kuno K., Lizasa H., Ohno S., Matsushima K.;			
RT	"The exon/intron organization and chromosomal mapping of the mouse			
RT	ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";			
RL	Genomics 46:466-471(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97150761; PubMed=8995297;			
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,			
RA	Matsushima K.;			
RT	"Molecular cloning of a gene encoding a new type of metalloproteinase-			
RT	disintegrin family protein with thrombospondin motifs as an			
RT	inflammation associated gene.";			
RL	J. Biol. Chem. 272:556-562(1997).			
RN	[3]			
RP	CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.			
RX	MEDLINE=99303657; PubMed=10373500;			
RA	Kuno K., Terashima Y., Matsushima K.;			
RT	"ADAMTS-1 is an active metalloproteinase associated with the			
RT	extracellular matrix.";			
RL	J. Biol. Chem. 274:18821-18826(1999).			
RN	[4]			
RP	FUNCTION.			
RX	MEDLINE=20389568; PubMed=10930576;			
RA	Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,			
RA	Ohno H., Matsushima K.;			
RT	"ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";			
RL	FEBS Lett. 478:241-245(2000).			
RN	[5]			
RP	FUNCTION, AND INDUCTION			
RX	MEDLINE=20243757; PubMed=10781075;			
RA	Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,			
RA	Richards J.S.;			
RT	"progesterone-regulated genes in the ovulation process: ADAMTS-1 and			
RT	cathepsin L proteases.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).			
CC	-1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE			
CC	INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY			
CC	SIMILARITY). ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH			

34	308	6.0	789	1	AD07_RAT	Q63180 rattus norv
35	297	5.8	480	1	DISA_TRIGA	P15503 trimeresuru
36	296	5.7	413	1	ACLA_AKAC	Q9pw35 agkistrodon
37	288.5	5.6	788	1	AD07_MOUSE	O35227 mus musculus
38	286	5.5	769	1	AD11_HUMAN	O75078 mus sapien
39	284	5.5	414	1	HRTD_CROAT	P15167 crotalus at
40	277.5	5.4	814	1	AD15_HUMAN	Q13444 homo sapien
41	277	5.4	773	1	AD11_MOUSE	O491v4 mus musculus
42	274.5	5.3	935	1	AD22_XENLA	O25956 xenopus lae
43	273.5	5.3	478	1	HRTD_CROAT	P34182 crotalus at
44	268	5.2	797	1	AD33_MOUSE	Q923w9 mus musculus
45	260.5	5.0	754	1	AD07_HUMAN	Q9h2u9 homo sapien

ALIGNMENTS

CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-I-LEU-1692
CC SITE. WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX.
CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC CELLS OF PREOVULATORY FOLLICLES.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FORIN ENDOPEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 7.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB001735; BAA24501.1; ALT_INIT.
CC EMBL; D67076; BAA11088.1; ALT_FRAME.
CC MEROPS; M12.222;
CC DR MGD; MGI:109249; Adamts1.
CC DR InterPro; IPR001762; Disintegrin.
CC DR InterPro; IPR002870; Pep_M12B_propep.
CC DR InterPro; IPR001590; Reprolysin.
CC DR InterPro; IPR000884; TSP1.
CC DR InterPro; IPR000130; Zn_MTpeptidse.
CC DR Pfam; PF00090; Tsp.1; 3.
CC DR Pfam; PF01421; Reprolysin; 1.
CC DR Pfam; PF01562; Rep_M12B_propep; 1.
CC DR SMART; SM00209; TSP1; 3.
CC DR PROSITE; PS00215; ADAM_MEPRO; 1.
CC DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC DR PROSITE; PS00092; TSP1; 3.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL; 1 48
FT PROPEP 49 253
FT CHAIN 254 968
FT SITE 206 206
FT METAL 402 402
FT ACT_SITE 403 403
FT METAL 406 406
FT METAL 412 412
FT DOMAIN 477 559
FT DOMAIN 560 617
FT DOMAIN 618 725
FT DOMAIN 726 850
FT DOMAIN 851 909
FT DOMAIN 910 968
FT DOMAIN 195 199
FT CARBOHYD 548 548
FT CARBOHYD 721 721
FT CARBOHYD 765 765
FT CARBOHYD 783 783
FT CARBOHYD 946 946
FT MUTAGEN 403 403
FT CONFLICT 335 335
FT CONFLICT 425 425
FT SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;

Query Match	48.1%	Score 2482.5;	DB 1;	Length 968;
Best Local Similarity	48.0%	Pred. No. 1.3e-167;		
Matches 478;	Conservative 158;	Mismatches 249;	Indels 111;	Gaps 22;
QY 1	M L L L G I L V L A F A G R T A G G - F E P E R E V V V P I R L P D I N G R R Y Y R G P - E D S G D O G L I F O I	57		
DB 37	L I L L A S I T M L L C A R G A H G R P T E E D E L V L P - S L E - - - - - R A P G H D S T T T R L - - R L	83		
QY 58	T A F Q E D F Y L H L T P D A O F L A P A F S T E H L G V P L Q G L T G G S - - - - - S D L R C F Y S G D V A	109		
DB 84	D A F G Q O L L H K L Q P D S G F L A G F T L Q V - - - - - G R S P G S E A Q H L D P T G L A H C F Y S G T V N G	138		
QY 110	E P D S F A A V S L C G L R G A F G Y R G A E V I S P L P N A S - - - - - A P A A Q R N S Q A - - - - - H L Q R R G	161		
DB 139	D P G S A A A L S L C E G V R G A F Y L Q G E E F F I Q P A G V A T E R L A P A V P E E S A R P Q F H I L R R R -	197		
QY 162	V P G G S G D P T S C G V A S G W N P A I L R A L D P Y K P R R A G F G E S R R R - - - - - - - - - - - - - -	206		
DB 198	- - - - - R G S G G A K G V M D -	243		
QY 207	- - - - - S G R A K R F V S I P R Y V E T L V V A D E S M Y K F H G A D L E H Y L L T L A T A A R L Y R H P S I L N P I	262		
DB 244	S G P G S I R K R F V S P R Y V E T L V A D Q S M A D F H G S G L K H Y L L T L F S V A A R F Y K H P S I R N S I	303		
QY 263	N I V V K V L L R D R D S G P K V T G N A A L T L R N F C A W K K L A K V S D K H P E Y W D T A I L F T R Q D L C	322		
DB 304	S L V V K I L Y I E Q K G P E V T S N A A L T L R N F C W K H S P S D R D P E H Y D T A I L F T R Q D L C	363		
QY 323	G A T T C D T L G M A D V G T M C D P K R S C V I E D D G L P S A F T T A H E L G H V F N M P H N V K Y C E E V F G	382		
DB 364	G S H F C D T L G M A D V G T V C D P S R S C S V I E D D G L Q A A F T T A H E L G H V F N M H D D A K H C A S L N G	423		
QY 383	K L R A N H M S P T L I Q I D R A N P W S A C S A A I T D F L D S H G D C L L D Q S K P L S L E D L P G A S Y	442		
DB 424	V T G D S H L M A S M L S L D H S Q P W S P C S A Y M V T S F L D N G H G E C L M D K P Q N P I K L F S D L P G T L Y	483		
QY 443	T L S Q O C E L A F G V G S K P C P - Y M Q Y C T K L W C T G R A K G O M V C Q T R H F P W A D G T S C G E G K L C L	501		
DB 484	D A N R Q C O F T F G E S K H C P D A A S T C T T L W C T G T S G G L L V C Q T K H F P W A D G T S C G E G K W C V S	543		
QY 502	G A C V E R H N L N K H - - - - - R V D G S W A K W D P G P C S R T C G G V Q L A R Q C T N P T P A N G K Y C E G V	558		
DB 544	K C V N K T D M - K H E A T P V H G S W G P W G D C S R T C G G V Q Y T R E C D N P V P K N G K Y C E G K	602		
QY 559	R V K Y R S C N L E P C S S A S G S K F R E E Q C E A F N G Y N H S T N R L T L A V A W P K Y S G V S P R D K C K L	618		
DB 603	R V R Y R S C N I E D C P D N - N G K T F R E E Q C E A H N E F S K A S F G N E P T V E W T P K Y A G Y S P D R C K L	661		
QY 619	I C R A N G T G Y F V L A P K V D G T L C S P D S T S Y C V Q G K I R A G C D G N L G S K K R F D K C G V C G G D	678		
DB 662	T C E A K G I G Y F V L Q P K V D G T P C S P D S T S Y C V Q G C V K A G C D R I D S K K K F D K C G V C G G N	721		
QY 679	N K S C K V T G L T K P M H G Y N F V A I P A G A S I D I R Q R G Y K G L I G D D N Y L A K N S Q G K Y L N	738		
DB 722	G S T C K M S G I V T S R P G Y H D I V T P A G A T N I E V K H R Q N R G S R N N G S F L A I R A A D G Y I L N	781		
QY 739	G H F V V S A V E R D L V V K S L R Y S G T G T A V E S I Q A S R P L E P L T V E L S V G K M T P P R V R Y S F	798		
DB 782	G N F T L S L E Q D L T Y K G T V L R Y S G S S A A L E I R S F S P L K E P L T I Q V L M V G H A L R P K I K F T Y	841		
QY 799	Y L P K E P R E D S K S H P K D P R G P S V L H N S V L S L S N Q V E Q P D D R P P A R W A G S W N G P C S A C G S G	858		
DB 842	F M K K T I E S - - - - - F N A I P T F S - - - - - E W V I E E W G E C S K T C G S G	874		
QY 859	L Q R A V P C R S A G O R T V P A C D A A H - - R P V E T Q A C G E - P C P T W E L S A W S P C S K S C G R G F O R	915		
DB 875	W Q R V V Q C R D I N G H - - - P A S E C A K E V K P A S T R P C A D L P C P H W Q V G D W S P C S K T C G K G I Y K	931		
QY 916	R S L K V C G G G R L L A R D O C N L H R K P O E - L D F C V L R P C	950		
DB 932	R T L K C V S H D G V L S N E S C D P L K K P K H Y I D F C T L T Q C	967		

aps 24;

16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
ADAMTS-1 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
ADAMTS1.
Rattus norvegicus (Rat).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
STRAIN=Sprague-Dawley; TISSUE=Brain;
Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
Little S.P.;
"Induction of a disintegrin and metalloproteinase with the
thrombospondin type I motif (ADAMTS).";
Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 18-967 FROM N.A.
STRAIN=Sprague-Dawley; TISSUE=Liver;
MEDLINE=20304099; PubMed=10847486;
Diamantis I., Luehl M., Hoesli M., Reichen J.;
"Cloning of the rat ADAMTS-1 gene and its down regulation in
endothelial cells in cirrhotic rats.";
Liver 20:165-172(2000).
-1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
SIMILARITY).
-1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684
SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
-1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
-1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (by similarity).
-1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CIRRHOTIC LIVER.
-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
-1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
SIMILARITY).
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
-1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
-1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; AF149118; AAC34012.1; -;
DR EMBL; AF304446; AAC29823.1; -;
DR MEROPS; M12.222; -;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTpeptidse.
DR Pfam; PF000090; tsp-1; 6
DR Pfam; PF01421; Reprolysin; 2
DR Pfam; PF01562; Pep_M12B_propep; 2.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS50215; ADAM_MPEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00427; DISINTEGRIN 1; FALSE NEG.
KW Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.

```
FT SIGNAL 1 54 POTENTIAL.
FT PROPEP 55 252 BY SIMILARITY.
FT CHAIN 253 967 ADAMS-1.
FT SITE 203 205 Cysteine switch (POTENTIAL).
FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 476 558 DISINTEGRIN-LIKE.
FT DOMAIN 559 615 TSP TYPE-1.
FT DOMAIN 616 724 CYS-RICH.
FT DOMAIN 725 857 SPACER.
FT DOMAIN 858 907 TSP TYPE-1 2.
FT DOMAIN 908 967 TSP TYPE-1 3.
FT DOMAIN 194 198 POLY-ARG.
FT CARBOHYD 547 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 720 740 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 21 31 I -> V (IN REF. 2).
FT CONFLICT 26 31 KFRSQ -> RSRGS (IN REF. 2).
FT CONFLICT 49 49 V -> A (IN REF. 2).
FT CONFLICT 72 72 R -> P (IN REF. 2).
FT CONFLICT 79 79 L -> TR (IN REF. 2).
FT CONFLICT 249 249 R -> G (IN REF. 2).
FT CONFLICT 262 265 TMLV -> NLLK (IN REF. 2).
FT CONFLICT 607 607 S -> F (IN REF. 2).
FT CONFLICT 936 936 L -> V (IN REF. 2).
FT CONFLICT 962 962 I -> T (IN REF. 2).
SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCDB4CF CRC64;
```

Query Match 47.9%; Score 2471; DB 1; Length 967;

Best Local Similarity 48.4%; Pred. No. 8.6e-167;

Matches 477; Conservative 161; Mismatches 256; Indels 92; Gaps 22;

```
QY 1 MILLGILTLAPAGRTAGG--FEPERVVPIRLDPDINGRRYWRGPDGSGOGLIFQIT 58
DB 37 LLLLSITMLLCVARGHGRPTDEELVL-----PSLERARGH-----DSTLLLRD 83
QY 59 AFOEDYVHLTPDQAFAPAFSTFHLGVPLQGLTGGS-----SDLRRCFYSGDVNAE 110
DB 84 AFGQQLHLKLQPDGSLAPGFLITQV-----GRSPGEAQHLDPGDLAHCFYGTGNGD 138
QY 111 PDFAVAVLGGURGAFGRGAEYVITSLPNAS-----APAAQNSQGA-----HLL--QRR 160
DB 139 PSSAAALSCEGVGAFYLGQEEFFIQAPAVATERLVPAEPKESSIAAPPREHILRRRR 198
QY 161 GVPGFSPD-----PTSRCGVASGNWPAILLRALDPYKPRRAGFGESRRSRRAKRV 214
DB 199 GSGGAKCGWDEETLPTSSGRESQNTPOWPLRNP--TPQAG--KPTPGSIRKKRV 254
QY 215 SIPRYVETLVADSMVKFPHGADLEHYLLTLLAARLYRHP-SILNPIVIVVKKVLLLRD 274
DB 255 SSPRYVETLVADQSMADFHGSLKHYLLTFSVAARFYKHP-SIRNSISLVVKKVILVYE 314
QY 275 RDSGPVGTGNAALTNFCAWOKKLNKVSCKHPEVWDFTAILFTRODLGCAITCDTLGMAD 334
DB 315 EQKGPEVTGNAALTNFNCFSWQKHNPSDDRDPEHYDTAILFTRODLGCHTCDTLGMAD 374
QY 335 VGTMCDDPKRSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHHMSTL 394
DB 375 VGTVCDDPSRCSVIEDDGLQAQAFATAHELGHVFNPHDDAKAFNGVSGDSHLMASML 434
QY 395 IQIDRANPWSAGSAALITDLSHGDCLLDQPSKPSISLPDLPASVTLSSOCELAGV 454
DB 435 SSLDHSOPWSPCSAYAVMTSFLDNGHCECLMDKQPNIPLSDPLGTLLDANRQCQFTGE 494
QY 455 GKPCP-YMOYCTKLWCTKAGQVMVQTRHPFPWADGTCGSGBKLCGLKAGCAVERHNLRKH 513
DB 495 ESTHCPDAASTCLWCTGTSGLLVCQTKHPFPWADGTCGSGBKWCVKCNKTKDM-KH 553
QY 514 ---RVDGSNAKWDPYPCSRCTCGGVQLARRQCTNPTTANGGKYCEGVRYKRSCLNLEPC 570
```

```
DB 554 FATPVHGWGMPGWDGCDSCRTCGGVQYTMRECDNPVPRNGGKYCEGKRVYRSCNIBDC 613
QY 571 PSSASGKSFREQCEAFNGYHNSTNRLTLAVAWVPKYSGVSPROKCKLICRANGTGYFYV 630
DB 614 PDN-NGKTFREQCEAHNEFSKASFGNEPTVETPKYAGVSPKORCKLTCEAKGIGYFFV 672
QY 631 LAPKVVDTCLSPDSTSVCVQKCLKAGCDNGLSKKRFKCGVCGGDNCKKVKVGLFT 690
DB 673 LQPKVVDGTPCSPDSTSVCVQKCYKAGCDRIIDSKKRFKCGVCGGNGSTCKKISGTVT 732
QY 691 KPMHGFNFWAIPAGASSIDIRQRYKGLIGDNDYALKNSQKYLNLNGHFWVSAVERDL 750
DB 733 STRPCYHIVIPAGATNIEVKHNRPRGRNNGSFLAIRAADGYIILNGNTLTSTLEODL 792
QY 751 VVKGLLYSGTGTAVESLQASRPILPTVEVLSVGKMTPPRVYSFLPK--EPREDK 808
DB 793 IYKGTVLRYSGSSAALIRSFSPKLEPTIQLVMVGHALRPKIKYTYFMKKTEP---- 848
QY 809 SSHPKDPRGSPVHLNLSVLSNQVQPDPRPARKVAGSWGPCSCASCSGSLQKRAYDCRG 868
DB 849 -----FNAITFS-----EWIEWEGECSKTCGSGWQRRVVECDR 883
QY 869 SAGQRTVPACDAAH--REVETQAGE--PCPTWELSAWPCSKSGRGFORSLKCVGHGG 925
DB 884 INGH---PASECAKEVKPASTRPCADLPCPRWQVQWDSWPCSKTCGKGYKKTAKLSHDG 940
QY 926 RLLARDQCNLHRKQOE-LDFCVLRPC 950
DB 941 GVLNSCEDPLKPKHYIDFCILTQC 966

RESULT 4
ATS8_HUMAN
ID AT88_HUMAN STANDARD; PRT; 890 AA.
AC Q9UPF79; Q9NZS0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2) (METH-8).
DE GN ADAMTS8 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=93367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S., Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity.";
RL J. Biol. Chem. 274:23349-23357(1999).
RN [2]
RP SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).
CC CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND KIDNEY.
CC CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
```


Genomics 62:312-315(1999).

CC - FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).

CC - COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC - SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

CC - TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.

CC - DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC - PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).

CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC - SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC - SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF175282; AAF25805.1; -.

DR HSSP: P34179; 1IAG.

DR MEROPS: M12.226; -.

DR MGD: MGI:1353468; Adamts8.

DR InterPro: IPR001762; Disintegrin.

DR InterPro: IPR001590; Repolysin.

DR InterPro: IPR000884; TSP1.

DR InterPro: IPR000130; Zn_Mrptptdse.

DR Pfam: PF00090; tsp_1; 2.

DR Pfam: PF01421; Repolysin; 1.

DR SMART: SM00209; TSP1; 2.

DR PROSITE: PS00142; ZINC_PROTEASE; 1.

DR PROSITE: PS02015; ADAM_MERO; 1.

DR PROSITE: PS00092; TSP1; 1.

DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.

DR Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat: Extracellular matrix; Heparin-binding.

FT SIGNAL 1 28 POTENTIAL.

FT PROPEP 29 228 BY SIMILARITY.

FT CHAIN 229 905

FT METAL 378 378 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 379 379 BY SIMILARITY.

FT METAL 382 382 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).

FT DOMAIN 453 541 DISINTEGRIN-LIKE.

FT DOMAIN 542 598 TSP TYPE-1 1.

FT DOMAIN 599 705 CYS-RICH.

FT DOMAIN 706 847 SPACER.

FT DOMAIN 848 905 TSP TYPE-1 2.

FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 905 AA; 98879 MW; 124D4132B33A0CAE CRC64;

Query Match 40.9%; Score 2113; DB 1; Length 905;

Best Local Similarity 45.5%; Pred. No. 1.5e-141;

Matches 431; Conservative 148; Mismatches 260; Indels 108; Gaps 24;

QY 2 LLLGLTL-----AFAGTAGGFEEVVPVIRLDPDINGRRYWRGPDSDGQGL 53

Db LLLLQLPPPLVCGAPAGPTGA--QASELVVPTL-----PGASE--L 55

QY 54 IFQITAFQEDYLLHTPDQAFLAFSTHGLGVPLQGLTGSSDLRCFYSGDGNAPDS 113

Db 56 AFLHSAFGQGVRLAPDASFLAEFEKIERLG-GSSAAAGGEPGLRGCFSTGVNGER 114

QY 114 FAASLCCGLRGAFGYRGAEEVVISPLPNASAPAAQRNSQA-----HLLQRGVGGP 166

Db 115 LAAMSCVAGSGSFLLAGEEFTIQ-----QGAGDSLDPHRLQRWG-PCQR 160

QY 167 SGDPSTSCGVASGNPAILRALDPYKPRRAFGEGRS-----RRSG----- 208

Db 161 REDP-----GLAAAEVFFLPQGLEWEVNGMGQOERSDEEDRQDKREGLKELKEDSRKV 216

QY 209 -----RAKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLATLARLHPSILN 260

Db 217 PPPGSTRKRFVSEARFVETLLVADASMAAFVGTDLQNHILTVMSAARIYKHSIRN 276

QY 261 PINIVVVKVLLLRDRDSGPKVTGNAALTNRNFCAMQKLNKVSFKHPEYWDTAILTRQD 320

Db 277 SVNLVVVKVILVEKERWGPVSDNGGLTLNFCNSWRRENKPSDRHPEHYDTAILTRQN 336

QY 321 LCG-ATTCDDILGMADVMTCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNVKCEE 379

Db 337 FCGGEQCDITGLMADVGTICDPKSCSVIKDEGLQRAYTTLAHELGHVLSMHPHDSKPCVR 396

QY 380 VFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDSHGDCLLDQSPKPSILPEDLPG 439

Db 397 LFGPMGYHMAFPFIHVNTLTPWSPCSAVYITELDDHGDCLLDAPTSVLPLPTGLPG 456

QY 440 AS--YTLUSQCELAFGVSGKPCP---YMQYCTKLWCTGKAKGMVCQTRH--FPWADGTS 492

Db 457 HSTLYELDOQCKQIFGPDFRHCNPNTSVEDICVOLCARHRDSDEPICHKNGSLWADGTP 516

QY 493 CGEGKLCIKGACVERHNLN--KRVDSGWAKWDPYGPCSRTCCGGVQLARQCTNTPAN 550

Db 517 CGPHLCIDGSCVLKEDVENPKAVVDGWDGWPWPGQCSRTCCGGIQTFSNRECDNMPQN 576

QY 551 GGYCEGVRYKRYSCNLEPCPSSASGSKSFREEQAFNGYHNHSTNRLTLAVAWPKYSGV 610

Db 577 GGRFCLGERVKYQSCNTECP--PNGSKSFEEQCEKYNAYNH--TDLGDFLOWVPKYSGV 633

QY 611 SPRDKCLICRANGTYFYVLAKPVVDGTLICSPDSTSVCGQKICRAGCDNGLSKRFD 670

Db 634 SPRDKCLFCRARGSEKVFVFEAKVIDGTLICGPDTLICVRGCVKAGCDHVVNSPKKLD 693

QY 671 KCGVCGDGNKCKVTGLFTKPMHGVNVAIPAGASSIDIRGKYLIGDDNYLAKN 730

Db 694 KCGVCGKGTACRKISGTFPFSFGINDIVTIPAGATNIDVKORSHPGVNDGSLYALKT 753

QY 731 SQKYLINGHFVSAVERDLVVKGLRYSGTGTAVESLQASRPILPTLVEVLSV-GKM 789

Db 754 ANGOYLINGNLATSAIBQDILVKGTILKYSGMTALERLQSFQALPELTVQLTVSGEV 813

QY 790 TPRVRYSFILPEPREDKSKHPDPRGSPVLSNLSVLSNQVEQDPPRPAKRWAGSWG 849

Db 814 FPPKRYTFFVPND--MDFSVQNSKERAT---NIQSL-----PSAEWVLGDWS 858

QY 850 PCSASGSGGLQKRAVDCRGSGAGQRTVPACDAHRPVTQACG-EPGP 895

Db 859 ECFSTCRGSGWRRTVECRDPGQAS-DTCDALKPEDAKPGSQPCP 904

RESULT 6

AT99 HUMAN STANDARD; PRT; 1629 AA.

ID AT99_HUMAN Q9NR29; AC Q9P2N4; Q9NR29; DT 16-OCT-2001 (Rel. 40, Created) DT 16-OCT-2001 (Rel. 40, Last sequence update) DT 15-JUN-2002 (Rel. 41, Last annotation update) DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9). GN ADAMTS9 OR KIA1312. OS Homo sapiens (human). OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. OX NCBI_TaxID=9606; RN [1] SEQUENCE FROM N.A. (SHORT ISOFORM). RP TISSUE=Fetal; RX MEDLINE=20396138; PubMed=10936055; Clark M.E., Kelner G.S., Turbeville L.A., Boyer A., Arden K.A.,

RA Naki R.A.;
RT "ADAMTS 9, a novel member of the ADAM-TS/Metallospandin gene
RT family.";
RL Genomics 67:343-350(2000).
RN [2]
RN SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RC MEDLINE=20181126; PubMed=10718198;
RX Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res 7:65-73(2000).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.
CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,
CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN
CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR
CC THYMUS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AF261918; AAF89106.1; -
DR EMBL: AB037733; BAA92550.1; -
DR HSSP: P15167.1ATL.
DR MEROPS: M12.021; -
DR Genew: HGNC:13202; ADAMTS9.
DR MIN: 605421; -
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_MTpeptcdse.
DR Pfam: PF00090; tsp_1; 11.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR SMART: SM00209; TSP1; 12.
DR PROSITE: PS00215; ADAM_MEPRO; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE: PS50092; TSP1; 9.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 287 BY SIMILARITY.
FT CHAIN 288 1629 ADAMTS-9.
FT DOMAIN 509 587 DISINTEGRIN-LIKE.
FT DOMAIN 589 642 TSP TYPE-1 1.
FT DOMAIN 645 752 CYS-RICH.
FT DOMAIN 753 880 SPACER.
FT DOMAIN 999 1053 TSP TYPE-1 2.
FT DOMAIN 1056 1108 TSP TYPE-1 3.
FT DOMAIN 1111 1156 TSP TYPE-1 4.
FT DOMAIN 1184 1239 TSP TYPE-1 5.
FT DOMAIN 1240 1295 TSP TYPE-1 6.

```
Db 694 GNTAYQLRDRVIDGTGCGQDNDICVQLCRQAGCDHVLNSKARRDKCGVCGDGNSSCK 753
QY 684 KYTGFTKPMHGYNFWAIPAGASIDIRQGRYKGLIGDDNYLALKNSQGYLNGHFVV 743
Db 754 TVAGFTNVHGYINTVVRIPAGANIDVRQHSFSGETDDDDNYLALSSKGFELNGFVV 813
QY 744 SAVERDLVVKGLSLRYSTGTAVESLQASRPILPTVEVLSVGMTPPRVRYFYLPKE 803
Db 814 TMAKREIRIGNAVVEYSGSETAVERINSTRIEQLLOLVSVGLNPDVRYFNFPIE 873
QY 804 PREDK---SSHPKDRGP-----SVLHNSVLSLNSQVEQPDPRPP--- 840
Db 874 DKPOQFYWNH-----GPWQACSKPCQGERKRLVCTRESQLTAVS---DQRCRLPQG 925
QY 841 -----ARVAGSWGPCSCSGSLQKRAVDCR-----GSAGQRTVPACDAHR 883
Db 926 HTEPCGTCDLRLWHVARSCECSQCGIGYRTLDIYCAKYSRLDGKTKKVDGFCSPHK 985
QY 884 PVETQACGECPT--WELSAWSPKSCGGRGFRSLKCVGHGGRLLARDQCNLHRKPQE 941
Db 986 PSNREKSCGECNTGWRYSAWTECSKSCDGGTQRRRAICVNTNRDVLDDSKCT-HQEKVT 1044
QY 942 LDFCVLRPC 950
Db 1045 IORCSEFFC 1053

RESULT 7
ATSA_HUMAN
ID ATSA4_HUMAN STANDARD; PRT; 837 AA.
AC 075173; Q9UN83;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1).
GN ADAMTS4 OR KIAA0688.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
RA Trzaskos J.M., Arner E.C.;
RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
RT family of proteins.";
RL Science 284:1664-1666(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;
RT "ADAMTS-4 genomic locus.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=10827174;
```

```
RA Tortorella M.D., Pratta M.A., Liu R.-O., Abbaszade I., Ross H.,
RA Burn T.C., Arner E.C.;
RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
RT aggrecan substrate recognition and cleavage.";
RL J. Biol. Chem. 275:25791-25797(2000).
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
CC ALZHEIMER'S DISEASE.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393
CC site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
CC -!- INDUCTION: BY INTERLEUKIN-1.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB014588; BAA31663.1; -
CC EMBL; AF148213; AAD41494.1; -
CC EMBL; AF044847; AAL02262.1; -
CC MEROPS; M12.221; -
CC Genew; HGNC:220; ADAMTS4.
CC MIM; 603876; -
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_Mtpeptdse.
CC Pfam; PF00090; tsp_1; 1.
CC Pfam; PF01421; Reprolysin; 1.
CC SMART; SM00209; TSP1; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS50215; ADAM_MEPRO; 1.
CC PROSITE; PS50092; TSP1; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; zymogen;
KW Extracellular matrix.
FT SIGNAL 1 51
FT PROPEP 52 212
FT CHAIN 213 837
FT SITE 194 194
FT METAL 361 361
FT ACT_SITE 362 362
FT METAL 365 365
FT METAL 371 371
FT DOMAIN 437 519
FT DOMAIN 520 576
FT DOMAIN 577 685
FT DOMAIN 686 837
FT DOMAIN 247 252
FT DOMAIN 68 68
FT CARBOHYD 77 77
FT CONFLICT 626 626
FT CONFLICT 682 682
FT CONFLICT 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;
SQ SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;
Query Match 37.1%; Score 1917; DB 1; Length 837;
```

Best Local Similarity 46.3%; Pred. No. 9,7e-128;

Matches 381; Conservative 123; Mismatches 267; Indels 52; Gaps 16;

```
QY 1 MLLGLTILAFAGTAGGEPREVVVPLRLDPDINGRYVWRGPDSDQGLIFQITAF 60
Db 37 LLLLLASLLPSARLASPLPREBIEVFPBKNGSVL-----PQSGAPAKLLRQLQAF 88
QY 61 QEDFYHLHTPDPAFLAPAFSTEHLGVPLQGLTGGSDLRRCFYSGDVNAEPDSFAAVSVC 120
Db 89 GETYLLLEODSGVQVEGLTVQVIGQAPE-LLGGAEP--GTVLTGTINGDPESVASLHWD 145
QY 121 GG-LRGAFGRAGYVISPPLNAPASAPAAQNSOGAHLORRGVPGPSDPTSRGCVASG 179
Db 146 GGALLGVQYRGAEHLQPLEGGTTPNSA--GGPGAHLRRK-----SPASQGGPMCNV--- 196
QY 180 WNPAILRALDPYKPRAGFSESRRSRGRKRFVSIPRYVETLVVVADESWMKVFHGADLE 239
Db 197 -----KAPLGSPSPRR--RAKRFASLSRFVELVVDKMAFHCAGLK 239
QY 240 HYLTLTATARYRHPSILNPINIVVVKVLLLRDSGPKVGTGNAALTLRNFCAWKKL 299
Db 240 RYLLTVMAAAKAFKHSIRNPVSLVTRVLVILGSGEGPQVGPSSAAQTLRSCAWQRLG 299
QY 300 NKYSDKHPEYWDTAILETRDILGAGTCTDLGNADVTGMDPKRSCSVTDEDDGLPSAFTT 359
Db 300 NTPEDSDPDHFDTAILETRDILGAGTCTDLGNADVTGMDPKRSCSVTDEDDGLQSAFTA 359
QY 360 AHELGHVFNPHNPVKNVCEEVFGKL-RANHMSPTLIQIDRANPWSACSAAIITDLDG 418
Db 360 AHELGHVFNHMLHDSKPCISLNGPLTSRHRVMAVHVDPEPWSPCSAREITDLDNG 419
QY 419 HGDCLLDQPSKPLPDLPGASVYLSQCELAFGVGSKPCPYM-QYCYKLVCTGKAKGQ 477
Db 420 YGHCLLDKPEAPLHLPVTFPGKDYADRCQQLTFGPDSDRHCPLPPPCAALWCSGHLNGH 479
QY 478 MVOCTRPEPWADGTSCEGKLCILKGACVERHNLNKHVD--GSWAKWDPYGPCSRTCGGG 535
Db 480 AMCQTKHSPWADGTPCPAOCNMGGRCLHMDQLQDFNIPOAGGWPMPGWDGSRCTCGGG 539
QY 536 VQLARRCQNTPTPANGKCYGEGVRKYRSCNLEPCPSASGKGFREPOCEAFNGYHNHSTN 595
Db 540 VQFSRDCRTPVRNGKCYGEGRRTRFRSCNTEDCP-TGSALTFRERQCAA--YVNHRTD 595
QY 596 RLTL---AVAWPKYSGVSPDKCLLCRANGTGYFVVLAPKVVDGTLGSPDSTSVCVQG 652
Db 596 LFKSPGPMWVRYTGYVAPQDQCLTCQARALGYVYVLEPRVVDGTPCSPDSSSVCVQG 655
QY 653 KCIRAGCNGNLGSKRDKCVCVCGDNKSKKVTGLFTKPMHGYNFVVAIPAGASSIDIR 712
Db 656 KCIHAGCDRIIGSKKFKDKCMVCGDGGSGSKQSGSEFKRYGYNVVTIPAGATHLVR 715
QY 713 QRGYKGLIGDDNYLALKNSQGYLLNGHFVVSAYERDLVKGSL-LRYSGTGTAVESLQA 771
Db 716 QQGNPG--HRSIYLALKLPDGSYALNGEYTLMPSTVDVLPVGAVSLRYSGATASETLSG 773
QY 772 SRPILPPLTVLSVGKMTTPRVRYSYFLPKPEPREDKSSHPKD 814
Db 774 HGPLAQPLTQVLVAGNPQDRLRYSFVFPVPRPTPTPTPDQ 816
```

RESULT 8

```
ATSS_MOUSE STANDARD; PRT; 930 AA.
ID ATSS_MOUSE
AC Q9R001;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (Implantin).
GN ADAMTS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9395124; PubMed=10464288;
RA Hurskainen T.L.; Hirohata S.; Seldin M.F.; Apte S.S.;
RT ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
zinc metalloproteases.
RL J. Biol. Chem. 274:25555-25563(1999).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (by similarity).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-
IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR
UNDETECTABLE LEVEL THEREAFTER.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF140673; AAD56356.1; -.
DR MEROPS: M12.225; -.
DR MGD: MGI:1346321; Adamts5.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSPI.
DR InterPro: IPR000130; Zn_MTPeptdse.
DR Pfam: PF00090; tsp_1; 2.
DR Pfam: PF01421; Reprolysin; 1.
DR SMART: SM00209; TSPI; 2.
DR PROSITE: PS50215; ADAM_MEPRO; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 261 POTENTIAL.
FT CHAIN 262 930 ADAMTS-5
FT SITE 209 209 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 411 411 BY SIMILARITY.
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 566 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE-1 1.
FT DOMAIN 624 731 SPACER.
FT DOMAIN 732 874 CYS-RICH.
FT DOMAIN 875 930 TSP TYPE-1 2.
FT DOMAIN 41 46 POLY-ALA.
FT DOMAIN 257 261 POLY-ARG.
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

16-OCT-2001 (Rel. 40, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2) (ADMP-2) (ADAM-TS 11).
ADAMTS5 OR ADMP2 OR ADAMTS11.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxId-9606;
[1]
RN SEQUENCE FROM N.A.
RC
TSISSUE=Liver;
MEDLINE=99367476; PubMed=10438522;
RA Abbaszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H., Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M., Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K., Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H., Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F., Arner E.C., Burn T.C.;
"Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family.";
RA J. Biol. Chem. 274:23443-23450(1999).
[2]
RN SEQUENCE FROM N.A.
RC MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien K., Lehmann R., Patterson D., Menzel U., Delabar J., Kumpf K., Lehmann R., Schudy A., Zimmermann W., Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E., Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P., Scharf M., Schoen O., Desario A., Reicheit J., Kauer G., Bloecker H., Ransier J., Beck A., Klages S., Hennig S., Riessmann L., Dagand E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.;
"The DNA sequence of human chromosome 21.";
RA Nature 405:311-319(2000).
[3]
RN SEQUENCE OF 413-930 FROM N.A.
RP
RC TISSUE=Fetal brain;
RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
"ADAM-TS, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases.";
RA J. Biol. Chem. 274:25555-25563(1999).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393 site.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE, CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN ARTHRITIC PATIENT.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC EMBL; AF142099; AAD49577.1; -
CC EMBL; AF001698; BAA95504.1; -
CC EMBL; AF001697; BAA95503.1; -
CC EMBL; AF141293; AAF02493.1; -
CC HSP; Q9PW35; 1BUD.
CC MEROPS; M12.225; -.
CC Genew; HGNC:221; ADAMTS5.
CC MIM; 605007; -.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSPL.
CC InterPro; IPR000130; Zn_MTPeptide.
CC Pfam; PF00090; tsp_1; 2.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSPL; 2.
CC PROSITE; PS00215; ADAM_MEPRO; 1.
CC PROSITE; PS00437; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS00092; TSPL; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 16 POTENTIAL.
FT PROPEP 17 261 POTENTIAL.
FT CHAIN 262 930 ADAMTS-5.
FT SITE 209 209 Cysteine SWITCH (POTENTIAL).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 411 411 BY SIMILARITY.
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 566 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE-1 1.
FT DOMAIN 624 731 SPACER.
FT DOMAIN 732 874 TSP TYPE-1 2.
FT DOMAIN 875 930 POLY-ALA.
FT DOMAIN 37 41 POLY-ARG.
FT DOMAIN 257 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 138 138 A -> G (IN REF. 2).
FT CONFLICT 614 614 R -> H (IN REF. 3).
FT CONFLICT 692 692 P -> L (IN REF. 2).
SQ SEQUENCE 930 AA; 101715 MW; B64281502F28193B CRO64;

Query Match 35.8%; Score 1847.5; DB 1; Length 930;
Best Local Similarity 44.0%; Pred. No. 8,9e-123;
Matches 392; Conservative 113; Mismatches 291; Indels 95; Gaps 22;

Qy 48 SGDQGLIFQITAFQEDFYHLTPDAQFAPAFSTEHGVLPLQLGTGSSDLR---RCFYS 104
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 82 SGGKGVLYVAGVRRLDLDERGSGVIAGF-----VPAGG--GTPAPWRHSHCFYR 133

Qy 105 GDVNAEDSPFANSLCGLRCAGFYCAEYVISP-----PNASAPAAQNSQGA----HLL 157
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 134 GTVDASPRSLAVFDLCGLDGLGFFAVKHARYTLKPLLRGPWAEKEKRGVYGDGSRILHY 193

Qy 158 QRRG-----VPGGPS-----GDPTSRGCVASGNPAIL--RALDPYKP 193
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 TREGFSEPALPRASCTPASTPEAHEHAPAHNSPNSGRAALAS-----QLDQSALSP--- 246

Qy 194 FRAGFES---RSRRSRGRKRFVSPRYVETLVVADESVMKFGADLEHYLTLTATAA 250
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 -AGSGPQTWRRRRRS-----ISRARQVELLVADASMARLYGRGLGYLLTLASIAN 299

Qy 251 RLYRHPSTLNPINIVVVKVLLLRDRDSDGPKVTGNAALTNRNFCNAWKLNKVDKHPY 310
```

```
Db 300 RYISHASINIRLAVKVVVLGDKDSLEVKNAATLKNFCKWQHQNOLGDDHEHY 359
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 311 DTAIFTRQDLCGATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMP 370
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 DAAILFTREDLGGHSCDTLGMADVGTICSPSCAVIEDDGLHAFTVAHEIGHLLGLS 419
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 371 HDNVKVCHEVFGKLRANHMSPFTLIQIDRANPWSACSNAITDLDSCGHGCLLDQSPK 430
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 HDDSKFEETFGSTEDKRLMSILTSIDASKPWSKTSATIFEFLDDGGHCLLDLPKK 479
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 431 ISLPEDLFGASVTLSSQOCELAFGSGKPCPYMQVCTKLVCTGKAKGQMVQCOTRHFPMADG 490
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 ILGPEELFGQYDATQQCNLTFGPEYSVCPGMDVCAWLCAVVRGQGMVCLTKLPAGEG 539
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 491 TSCGEGKILCKGACVER-----HNLNKHVRDGSNAKMDPYGPCSRTCGGGVOLARRCTN 545
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 540 TPCGKGRICLOGKVDKTKKKYVSTSSH---GNWGSWGSWGQCSRSCGGGVQFAYRHNN 596
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 546 PTPANGKYGCEGVRYKRSNLEPCPSSASGSKSPREECEAFNGYNHSTNRLTLAVAWVP 605
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 597 PAPERNGRYCTGKRAIYRSCSLMPCP--PNGKSPRHEQCEAKNGYQSDAKGVKTFVEWVP 654
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 606 KYSGVSPRDKCKLICRANGTGYVVLAPKVVVDGTLCSDDSTSVCGQKICAKGCDGNLGS 665
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 655 KYAGVLPADVCKLTCKRAGTGYVVFSPKVTGTECRYSNSVCRGKCVGTGCGGIIGS 714
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 666 KKRFKCGVCGGDNKCKKVTGLFTKPMHGNFVVAIPAGASSIDIRQRYKGLIGDDNY 725
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 KLIQDKCGVCGGDNSSCTKIYGFNKKSKGYTDVVRIPEGATHIKVRQFKAQDQTRFTAY 774
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 726 LALKNSQKYLNGHFVVSVERDLVVGSLRLRYSGTGTAVESL-----QASRPILPLT 780
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 775 LALKKNGEYLNGKYMISTSETIIDINGTVMNSYSGWSHRDDFLHGMGYSATKEI---LI 831
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 781 VEVLSVGMKTPPRVYSFVLPKPEKREDKSSHPKDPGRPSVLHNSVLSL-SNQVEQPDPR 839
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 832 VOILAIDPTKPLDVYSFVFPK-----STPK-----VNSVTSHGSKVGSHTSQP 877
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 840 PARWAGSVPSCSASCGLQKRVDCRGSAQORTVPACDAHRAHVPETQAC 890
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 878 --QWVTGPMACSRCTDGTGWHTRTVQCO--DGNRKLAKGCLPSQRPSPAFKQC 925
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
ATSA4_RAT STANDARD; PRT; 630 AA.
ID ATSA4_RAT
AC Q9ESP7; Q9ESP8; Q9ESP6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-4 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1) (Fragment).
DE (Fragment).
DE ADAMTS4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=WiStar; TISSUE=Brain;
RX MEDLINE=20415831; PubMed=10961658;
RA Sato H K., Suzuki N., Yokota H.;
RT "ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin motifs) is transcriptionally induced in beta-amyloid treated rat astrocytes.";
RL Neurosci. Lett. 289:177-180(2000).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393 site.
```

```

242  COLTFEGDPDSHCPOLP P P P C A A L W C F G H L N G H A M C O T K H S P W A D G T P C G P A Q A C W G R C L H      301
507  R H N L N K H R V D - G S W A K W D Y G P C S T C G G V Q L A R O C T T P T P A N G K Y C E G V R Y K Y R S      564
302  V D O L K F N I P O A G W G P W G M G C D K C T C G G V Q F S R D C T K P V P R N G G K Y C E G R H T P F R S      361
565  C N L E P C P S S A S G K F R E O C E A F N G N H S T N R L T L - - - A V A W P V K Y S G V S P R D K C K L I C R      621
362  C N T K N C P - H G S A L T F R E E Q C A A - - - Y N H R T D L F K S P G P M D W V P R Y T G V A P R D O C K L T C Q      417
622  A N G T G Y F V L A P K V D G T L C S P D S T S V C V G Q C I R A G C D N G L S K K R F D K C G V C G G D N K S      681
418  A R A L G Y Y V L E F R V A D E T P C S P D S S V C V G R C I H A G C D R I I G S K K R F D K C M W C G G N G S S      477
682  C K K V T G L F T K M H G Y N F W A I P A C A S S I D I R Q R G Y K G L I G D D N Y L A L K N S O G K Y L L N G H F      741
478  C S K Q S G S F K F P R Y G S D W V T I P A G R T H I L V R O G G S G L - - K S I Y L A L K L A D G S V A L N G E Y      535
742  V Y S A V E R D L V V K G S L - L R Y S G T G T A V E S L Q A S R P I L E T V L S V G K M T P P R Y R Y S F Y L      800
536  T L M P S T D V L P G A V S I R Y G R T A A S E T L S G H G L P A Q P L T L Q V L V A G N P Q N V R L Y S F F V      595
801  P K E P R E D K S H P K D P R G P      818
596  P R P V - - - - - P S T P R P P      606

RESULT 11
ID      AT10_HUMAN      STANDARD;      PRT;      1077      AA.
AC      Q9H324;
DC      15-JUN-2002 (Rel. 41, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE      with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
DE      ADAMTS10.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Apte S. S.;
RT      "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT      thrombospondin type I repeats.";
RL      Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC      -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC      matrix (By similarity).
CC      -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC      FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
CC      similarity).
CC      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC      -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC      -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC      -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AF163762; AAG35563.1; -
CC      DR      MEROPOS; M12.235; -
CC      DR      Genew; HGNC:13201; ADAMTS10.
CC      DR      InterPro; IPR001762; Disintegrin.
CC      DR      InterPro; IPR002870; Pep_M12B_propep.
CC      DR      InterPro; IPR001590; Reprolysin.
CC      DR      InterPro; IPR000884; TSPI.
CC      DR      InterPro; IPR000130; Zn_MTpeptidase.

```

DR Pfam: PF00090; tsp_1; 5.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR SMART: SM00209; TSP1; 5.
DR PROSITE: PS00215; ADAM_MEPRO; 1.
DR PROSITE: PS00092; TSP1; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT NON_TER 1
FT PROPEP <1 207 BY SIMILARITY.
FT CHAIN 208 1077 ADAMTS-10.
FT METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 367 367 BY SIMILARITY.
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 CYS-RICH.
FT DOMAIN 680 802 SPACER.
FT DOMAIN 521 577 TSP TYPE-1 1.
FT DOMAIN 799 860 TSP TYPE-1 2.
FT DOMAIN 862 918 TSP TYPE-1 3.
FT DOMAIN 922 976 TSP TYPE-1 4.
FT DOMAIN 981 1031 TSP TYPE-1 5.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 769 769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCB9F587 CRC64;

Query Match 25.18; Score 1296; DB 1; Length 1077;
Best Local Similarity 35.48; Pred. No. 9.4e-94;
Matches 335; Conservative 112; Mismatches 376; Indels 123; Gaps 32;

QY 24 EVVVPRLDPPINGER-----RYTWGPEDSGDGLIFQITAFQEDFYHLTPDAQF 74
DB 14 EIAFPTRVHD--NGALLAFSPPPRRQRGTGATAESRLFYKVASPSTHELLNLTSSRL 71

QY 75 LAPAFSTHGLVPLQGLTGGSSDLRCFYSDVNAE--PDFAAVSLCGGLRGAFYGAEE 133
DB 72 LAGHVSVEYW--TREGLAWQRAARPHCLYAGHLQAGQSSSHVAISTCGGLHGLIVADEE 129

QY 134 YVISPLNASAPAAQRNSOHAHLQRRGVPCGPGDPTSRGV--ASGW--NPAILRALD 189
DB 130 YLIEPLHGGPGKSPESGPHVYKSSLRPHLD--TAGGVDEKPKRPMWLTIX 187

QY 190 PYKPRRAGGESRRSRRGRKFRVPIRYVETLIVADESMVKPHG--ADLEHYLLTLAT 248
DB 188 P--PPARPLGNETERGQPG--LKRSVSRERYVETLIVADKMMVAYHGRDRVQYVLAINI 244

QY 249 ARLYRHPISILNPINIVVVKVLLLRDSDGPKVTGNALTLRNCAMQKL-----NK 301
DB 245 VAKLFQDSLSTGNILVRLILITLTDQPTLEITHHAGKSLDSFCWKQKSVNHSHGNA 304

QY 302 VSDKHPEYWDYALFTTRDLC--GATTCDTLGMADVGTMCDPKRSCSVIEDDGLPSAFT 359
DB 305 IPENVANHDYAVLTIRYDICIYKPKCGTLGLAPVGMGERSCSVNEDIGLPQAFIT 364

QY 360 AHELGHVNMHDNVKVEEVEFG-----KLRANHMMSPTLIIDRANP--WSACSAALITD 413
DB 365 AHEIGHTFGMNHGVDGNSCGARGQDPAPKMAAHITMTK-----NPFVWSSCNRDITS 417

QY 414 FLDSDGCLLDLPQSK--PISLPEDLPASLYTLSOCELAFGVGSKPCPYMOCYKFLWCTG 472
DB 418 FLDSDGLGLNRRPPQDFVPTVAPGQAYDADECQFHQGVKSKQCKYGEVCSLWCLS 477

QY 473 KAKGMVQCTRHFPWADGTSCEGKL-----CLKGACVYERHNLKRVHDSWAKWDYGP 528
DB 478 KSNR---CITNSIPAAEGTLQTHITDKGWYKRVCP--FGSRPEGVGDGANGPWTWGD 533

QY 529 SRTCGGVQLARRQCTNPTPANGKYCEGVRYVKYRSCNLEPCPSASGSKSPREOCFAFN 588
DB 534 SRTCGGVYSSSRHSDSPRTIGGKYCLGERRRHRSCNTDDCPGGS--QDFREVQCSEFD 591

QY 589 G-----YNHSTNRLTLAVAWPKYSGVSPDKCKLLCRANGTCYFVVLAPKVVDTGLC 641
DB 592 SIPFRGKFKYKWKTYR-----GGGVK---ACSLTSLAEGFNFYTERAAVVDGTPC 638

QY 642 SPDSTVSCVOGKCIKAGCDNGLGKRFDKGVCVGGDNKSKKVTGLFT--KPMHGYNFV 699
DB 639 RPDVTDICVSGECKHVGCDDRLGDLREDKRCVCGDGSACETIEGVSPASGAGYEDV 698

QY 700 VAIPAGASSIDIRQRYKGLIGDDNYLALKNSQGYLLNGHFVWSAVERDLVWKSLLRY 759
DB 699 VNIKGSVHIFIQD-----LNLSSLHALKGDQESLLEGLPGTQPQHR-LPLAGTTTQL 752

QY 760 STGTAVESLQASRIPLILEPLTVEVLSVGMKTPPVRYSYFLPKPEPKEDKSSHPRGPGS 819
DB 753 RQGPDOVQSLAELGPINASLIYVNLARTEL--PALRYRNPAP----- 792

QY 820 VLHNSVLSLSNOVEQPDPRPRAWVAGSGPCASCGSLQKRAVDCRGSAGQRTV-PAC 878
DB 793 -----IARDSLPYSHYAPWKCSAQACAGSQVQAVECRNQLDSSAVAPHY 839

QY 879 DAAHR--PVETOAGC-EPC-PTWELSAWSPCSKSCGRGFQRRSLKC 920
DB 840 CSAHSLPKRQRACNTEPCPPDWVVGNNSLCSRSRCDAGVRSRVVC 885

RESULT 12
ID AT12_HUMAN STANDARD; PRT; 1593 AA.
AC P58397;
DC 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S.; Argueelles J.M.; Fernandez P.L.; Lopez-Otin C.;
RT ADAM-TS12, a novel human disintegrin, and intracellular processing of
RT organization involving multiple thrombospondin-1 repeats.;
RL J. Biol. Chem. 276:17932-17940(2001).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -1- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; AJ250725; CAC20419.1; -	
DR	Genew; HGNC:14605; ADAMTS12.	
DR	MIM; 606184; -	
DR	InterPro; IPR001762; Disintegrin.	
DR	InterPro; IPR002870; Pep_M12B_propep.	
DR	InterPro; IPR001590; Reprolysin.	
DR	InterPro; IPR000884; TSPl.	
DR	InterPro; IPR000130; zn_MTPeptide.	
DR	Pfam; PF000090; tsp_1; 6.	
DR	Pfam; PF01421; Reprolysin; 1.	
DR	Pfam; PF01562; Pep_M12B_propep; 1.	
DR	SMART; SM00209; TSPl; 8	
DR	PROSITE; PS02115; ADAM_MEPRO; 1.	
DR	PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.	
DR	PROSITE; PS50092; TSPl; 2.	
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.	
KW	Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;	
Repeat;	Extracellular matrix.	
FT	SIGNAL 1 25	POTENTIAL.
FT	PROPEP 26 240	BY SIMILARITY.
FT	CHAIN 241 1593	ADAMTS-12.
FT	DOMAIN 465 544	DISINTEGRIN-LIKE.
FT	DOMAIN 545 596	TSP TYPE-1 1.
FT	DOMAIN 597 700	CYS-RICH.
FT	DOMAIN 701 826	SPACER 1.
FT	DOMAIN 827 881	TSP TYPE-1 2.
FT	DOMAIN 886 943	TSP TYPE-1 3.
FT	DOMAIN 947 995	TSP TYPE-1 4.
FT	DOMAIN 996 1315	SPACER 2.
FT	DOMAIN 1316 1364	TSP TYPE-1 5.
FT	DOMAIN 1367 1423	TSP TYPE-1 6.
FT	DOMAIN 1426 1471	TSP TYPE-1 7.
FT	DOMAIN 1426 1471	TSP TYPE-1 8.
FT	DOMAIN 302 305	POLY-GLU.
FT	SITE 208 208	CYSTEINE SWITCH (POTENTIAL).
FT	METAL 392 392	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE 393 393	BY SIMILARITY.
FT	METAL 396 396	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL 402 402	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD 105 105	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 125 125	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 215 215	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 485 485	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 685 685	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 790 790	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 951 951	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1104 1104	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1275 1275	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1300 1300	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1320 1320	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1371 1371	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1378 1378	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1503 1503	N-LINKED (GLCNAC. .) (POTENTIAL).
DR	SEQUENCE 1593 AA; 177545 MW; 07F9F48B63BD83A3 CRC64;	
DR	SEQUENCE	

<hr/>					
Query Match	24.7%	Score 1275;	DB 1;	Length 1593;	
Best Local Similarity	32.4%;	Pred. No. 4.6e-82;			
Matches 330; Conservative 155;	Mismatches 374;	Indels 158;	Gaps 37;		
<hr/>					
QY	21	PEREVVPIRLD-----PDINGRRYYRGPEDSGDGLFIQTAFQEDFYHLH 68			
		: :	:	:	:
Db	50	PEYHVGPVRVDASGHETSYGLHYPTTSRR---KRDLDGSEDWYYSRISHEEKDLFFNL 106			
		:	:	:	:
QY	69	TPDAOFAPAFSTEHLGVPLQGLTGSSDLRCFFSGDVNAEPD--SFAAIVSLCGGLRGA 126			
		: : :	:	:	:
db	107	TYNQGFLSNSIYMEKRYGNLSHVKKMAASAPLCHUSGITVLQOQTRVGTAALSAACHGLTGF 166			
		: : :	:	:	:
<hr/>					
RESULT 13					
		ID_ATS2_HUMAN	STANDARD;		PRT; 1211 AA.
		AC O95450;			
		DT 16-OCT-2001 (Rel. 40, Created)			
		DT 16-OCT-2001 (Rel. 40, Last sequence update)			
		DT 15-JUN-2002 (Rel. 41, Last annotation update)			
		DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)			
		DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)			
		DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I			
		DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI)			

DE (Procollagen I/II amino-propeptide processing enzyme).

GN ADAMTS2 OR PCNP OR PCNP1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.

RC TISSUE=Skin;

RX MEDLINE=99347935; PubMed=10417273;

RA Collige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,

RA Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,

RA Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.;

RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis

RT are caused by mutations in the procollagen I N-proteinase gene.";

RL Am. J. Hum. Genet. 65:308-317(1999).

CC -!- FUNCTION: Cleaves the propeptides of type I and II collagen prior

CC to fibril assembly. Does not act on type III collagen. May also

CC play a role in development that is independent of its role in

CC collagen biosynthesis.

CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain

CC alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(1) chains

CC at Ala-1-Gln.

CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO

CC COLLAGEN TYPE XIV (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular

CC matrix (By similarity).

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: LPNPI (SHOWN HERE) AND SPNPI;

CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-

CC PROCOLLAGEN PEPTIDASE ACTIVITY.

CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON

CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.

CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT

CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY

CC SIMILARITY).

CC -!- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos

CC syndrome type VIIC (EDS-VIIC), a recessively inherited connective-

CC tissue disorder characterized clinically by severe skin fragility

CC and joint hypermobility and biochemically by the presence in skin

CC of procollagen incompletely processed at the N-terminus.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.

CC -!- CAUTION: HAS SOME TIMES BEEN REFERRED TO ADAMTS3.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL: AJ003125; CA05880.1; -

DR MEROPS: M12.301; -

DR Genew: HGNC:218; ADAMTS2.

DR MIM: 604539; -

DR MIM: 225410; -

DR InterPro: IPR001762; Disintegrin.

DR InterPro: IPR002870; Pep_M12B_propep.

DR InterPro: IPR001590; Reprolysin.

DR InterPro: IPR000884; TSP1.

DR InterPro: IPR000130; zn_MTPeptdse.

DR Pfam: PF01421; Reprolysin; 1.

DR Pfam: PF01562; Pep_M12B_propep; 1.

DR SMART: SM00209; TSP1; 4.

DR PROSITE: PS00215; ADAM_MEPRO; 1.

DR PROSITE: PS00092; TSP1; 1.

DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.

DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;

KW Alternative splicing; Ehlers-Danlos syndrome.

FT SIGNAL 1 29

FT PROPEP 30 253

FT CHAIN 254 1211

FT METAL 408 408

FT ACT_SITE 409 409

FT METAL 412 412

FT METAL 418 418

FT DOMAIN 480 560

FT DOMAIN 561 617

FT DOMAIN 618 722

FT SITE 691 693

FT DOMAIN 723 851

FT DOMAIN 852 911

FT DOMAIN 912 974

FT DOMAIN 975 1030

FT DOMAIN 40 43

FT DOMAIN 185 188

FT CARBOHYD 112 112

FT CARBOHYD 251 251

FT CARBOHYD 949 949

FT CARBOHYD 993 993

FT CARBOHYD 1031 1031

FT CARBOHYD 1098 1098

FT CARBOHYD 1145 1145

FT CARBOHYD 1150 1150

FT VARSPLIC 544 566

FT VARSPLIC 567 1211

FT SEQUENCE 1211 AA; 134722 MW; BECEE25C23CAD2D CRC64;

SQ

Query Match 23.8%; Score 1230; DB 1; Length 1211;

Best Local Similarity 33.1%; Pred. No. 5e-79;

Matches 347; Conservative 134; Mismatches 404; Indels 162; Gaps 47;

Qy 9 LAFAGRTAG---GPEPREVVVPIRLDPD-----INGRR-----YVW 42

Db 39 LAAADPPGPGHGAERILAVPVRTDAQGRVLSHVVSAAATSRAGVRARAAVPTSPFP 98

Qy 43 RGPEDSGDGLIFQITAFQEDFYLRHLPDAQFLAPAFSTEHLGVLPLOGLTGGSDLRRCF 102

Db 99 GGNREEPGSHLFYVNTVGRDLHLRLPRLNARLVAPGATMEWQGE--KSTTVPEVLGSL 156

Qy 103 YSGDVN--AEPDSPAASVLCGLRGAFYGAEGYVISPLPNASAPAAQRNSQG-AHLQR 159

Db 157 YVGDVAGLAEASS-VALSNCDGLAGLIRMEEEFFIEPLEKGL--AAQEAEGRVHVYR 213

Qy 160 RGVPGGSGDPTSRGVSAGSNPAILLRALDPYKPRRAGFSGESRRSRRAKRVSPY 219

Db 214 RPPTSPPLGGPQA---LDTG---ASLDSLSLS-RALGVLEEHANSRRRRARHAADDY 266

Qy 220 -VETLVVADESVMKAFHGAD-LEHYLLTLTLLATAARLYRHSILNPINIVVVKVLLLRDSD 277

Db 267 NIEVLGVDDSVVQFHGKGVKQVLLTLMNIVNEIYHDESLGAHINVVLRIILSYGKS 326

Qy 278 GPKV-TGNAALTNRNFCAMQKLNKVSOKHPYWDFTAILFTRQDLCGATCTDLGMADVG 336

Db 327 MSLIEIGNPSQSLNVCWYWAYLOOQKPDGTGHDHEYHDHAIFTLRDGF-GPSGMQ--GYAPVT 383

Qy 337 TMCDPKRSCTVEDDGLPSAFTTAHELGHVFNMPHDNV-KVCEEVFKLRANHMSPITLI 395

Db 384 GMCHPVRSCITLNHEDGSSAFVVAHETGHVLMGHEDDGGNRCGD---EVLGSTMALPVQ 440

Qy 396 QIDRANPWSACSAIITDPLDSDHGCDLLDQP---SKPISLPEDPLGASVTLSSQCEELAF 452

Db 441 AAFHREHWSRCSQQLSRYLHS--YDCLDDPFAHDNP-ALPQ-LPGLHYSMNEQCRDF 496

Qy 453 GVGSKPCPYMQY---CTKLWCTGRAGKQWVCQTRHFPWADTSGEGKCLCKAGCV---- 505

Db 497 GLGYMCMCTAFRTDFDCKQLWCS-HPDNPYFCKTKKGGPPLDGTMCAPKHKCFKGIWLT 555

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

DR EMBL; AF140675; AAD56358.1; -.
DR HSSP; P15167; IATL.
DR MEROPS; M12.231; -.
DR Genew; HGNC:223; ADAMTS7.
DR MTM; 605009; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS0215; ADAM_MEPROSE; 1.
DR PROSITE; PS50092; TSPI; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
FT REPEAT; Extracellular matrix.
FT SIGNAL 1 27 POTENTIAL.
FT FT PROPEP 28 232 BY SIMILARITY.
FT FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 BY SIMILARITY.
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match          35.5%;   Score 1212.5; DB 1; Length 997;
Best Local Similarity 35.1%;   Pred. No. 6.8e-78;
Matches 353; Conservative 116; Mismatches 377; Indels 159; Gaps 46;

QY 1 MLLGILTLPAGRTAGGFEFER---EVVVPRLDP-----DINGRR- 39
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 18 LLLLALAPGAPGAPGRATEGRAALDIVHPVDAGGSFLSYELWPRALKRKRVSRRD 77
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 40 ---YY---WRGPESGDGLIFQTAFQDELYHLHTPAQFLAPAF--STHGLGVPLQGL 91
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 78 APAFYELQYRGRE-----LRFNLTANQ-----HL----LAPGVSETRRRG----GL 116
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 92 TGGSSDLR-----RCFYSGDVNAEPD---SFAAVSLCGGLRGAFGYRGAEYISPLPNASA 144
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 117 --GRAHIRAHTPACHLLGEVQ-DPELEGLLAISACDKLGVFQLSNEDYFIEPLD--SA 171
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 145 PAARNSSQGAHLQRQRYVP-----GGPSGDTPSRCGVASGNWPAILRALDPYKPPRRAGFG 199
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 172 PARPCHAQ-PHWYVKQAPERLAORGDSSAP-STCGV-----QVYPELESRRRW 219
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 200 ESRRRSRGRAKRF-----VSIPRVETVLVADESVMKFHG-ADLEHYLLTLTATAARLYR 254
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 220 EQROQRRLRLRHORSVSKVEYTVLWADAKMVEFHGQPQVESYVLTINNMYAGLFPH 279
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 255 HPSILNPNIVVKKVLLLRDSDGPKVTGNAALTURNFCAWCKLNKYSDKHPYEWDTAI 314
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 280 DPSIGNPHTITVRVLLEDEEEDLKITHHADNTLKSCKWKQSINMKGDAPHLDHTAI 339
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 315 LFTRDOLCGATT--CDTLGMADVGTMCDDPKKSCSVIEDGLPFSAPTAAHELGHVENMPHD 372
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```

Db 340 LLTRKDLCAAMNRCEFTLGLSHVAGMCQPHRSQCSINEDTGLPLAFTVAHELGHSGFIQHD 399
Qy 373 -NVKCVEEVFGKLRANHHMPTLLIQIDRANPWSACSAAIITDFDLSGHGDCLLDQPSKP- 430
Db 400 GSGNDCEPV-GK--RPTIMSPQLLYDAAPLTWSRCSRQYITRELDRGWGLCLDDPPAKDI 456
Qy 431 ISLPEDLPGASYYTLLSOCELAFGVGSKPCPYM-QYCTKLWCTGKAKGQWVCQTRHFPWAD 489
Db 457 IDFSVPVGVLYDVSHQCRLOYGAYSACFEDMDNVCHTLCWCSVGT----TCHSKLDAAYD 512
Qy 490 GTSCGEGKCLIKGACVERHNLKRVHVDGWSNAKWDYPCPSRSTCGGVQQLARRCTNPTPA 549
Db 513 GTRCGENKWLCSGCEVP-VGFRPEAVDGGWSGWSANSICSRSCGMGVQSAERCTQPTPK 571
Qy 550 NGGKYCGVRVVKYRSCNLEPCPSPASAGSKSPREQCCEAFNGYHNHSTNRLTLAVAVWPYSG 609
Db 572 YKGRYCVGERKRFRLCNLQACP--AGRPSFRHVQCSEHFDAMLYKGQLHT----WVPVWD 625
Qy 610 VSPRDKCKLICRANGTGYFVLAPKVVVDGLPCSP--DSTSVCVQGGKCIKAGCDGNLGSKK 667
Db 626 VNP---CELHCRPANEFYAKKLDAVVDGTPCYQVRSRDLCLINGLCKNVGCFEIDSGA 682
Qy 668 RFDKCVGGGDNKSKVKVTLGTFKPMH-GYNFVVAIPAGASSIDIRQGYKGLIGDNDYL 726
Db 683 MEDRCGYCHNGSTCTHTVSGTFFEEAGLGVVDVGLIPAGAREIRIQE-----VAEAAFL 737
Qy 727 ALKNSQ-GKYLNGHFVVSVERDLVVKGLLRYSGTGTAVESLQASRILPLTVEVL- 784
Db 738 ALRSEDEKFLNGWTIQ-WNGDYQVAGTFTFYARGN-WENLTSPPGKTPGVWIQVPA 795
Qy 785 -----SVGKMTPPRVKYSFYLPKEPRDKSSHPRGSPVYHNHNSVLSNQVEQPD 837
Db 796 SRGPGGSRGSGVPRPSTLHGSRSPGGVSPGSVTEPGSEPGPPAAASTSVS----- 845
Qy 838 RPPARW---VA-----GSWGPCASCSGSLQKRAVDCRGs-----AGQTVVPACDAHRP 884
Db 846 -PSLKWPNLVAAVHRGWG--QAPLGLGWRHRLVLMGPRLPQLTQLFQESNPGVHYEY-T 901
Qy 885 VETQACGE-----PCPTWELSAWSPCSKSCGRG--FORRSLKCVG 922
Db 902 IHREAGGHDEVPPVPSVSWHYGPWTKCTVICGRGEKWRHSPTCRG 946

RESULT 15
ID AT52_BOVIN STANDARD; PRT; 1205 AA.
AC P79331;
CT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-2 precursor (BC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endoropeptidase) (pNPI).
DE ADAMTS2 OR NPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=97225960; Pubmed=9122202;
RA Collige A., Li S.W., Sieron A.L., Nussgens B.V., Prockop D.J.,
RA Lapierre C.M.;
RT "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT new member of the superfamily of zinc-metalloproteinases with binding
RT sites for cells and other matrix components.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RP [2]
RP PARTIAL SEQUENCE.

```

```
FT DOMAIN 846 905 TSP TYPE-1 2.
FT DOMAIN 906 968 TSP TYPE-1 3.
FT DOMAIN 969 1024 TSP TYPE-1 4.
FT SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 31 35 POLY-ALA.
FT DOMAIN 177 180 POLY-GLU.
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 943 943 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 987 987 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;

Query Match 23.18; Score 1192; DB 1; Length 1205;
Best Local Similarity 31.78; Pred. No. 2.4e-76;
Matches 329; Conservative 134; Mismatches 403; Indels 172; Gaps 42;

QY 1 MLLGLITLAFAGTAG-----GFEPEREVVPRIID----- 32
DB 16 LLLLLLLPADARLAAAADPPGPGQGHGAERILAVFVRTDAQRLVSHVSAATAPAG 75
QY 33 -----PDINGRYYWRGPDSDGGLIFQITAFQEDFYHLTPDAQFLAPAFST 81
DB 76 VTRRAAPAIPLSG-----GSEEDPGRLFYNTVFGRLRLRPNARLAVPAGATV 129
QY 82 EHLGVPLQGLTGSSDLRCFYSGDV-NAEPDSFAAVSLCGGLRGAFGYRGAEYVISPLP 140
DB 130 EQWGE--SGATRVPLLTGLVGVAGLAESSVALSNCDSLGLAGLRMEEEFFIEPLE 187
QY 141 NASAPAAQRNSOG-AHLQRGVP-----GGPSGDTSRGCVASGWNPAIRALDPYKP 193
DB 188 KGL--AAKEAEGRVHYVYHRTTSRPPPLGQPQALDT---GISADSLDSLAL----- 237
QY 194 RRAFGESRRRSRRAKRFVSIPIRY-VETLVVADESVMKFGHAD-LEHYLLTLATAAR 251
DB 238 --GVLEERVNSRRMRHAADDYNIIEVLGVDDSVVQPHGTEHVQKYLTLNMVNE 294
QY 252 LYRHP SILPNIYVVKVLLLRDRSGPKV-TGNALTLRNFCAWOKKLNKVDKHPYEW 310
DB 295 IYHDESLAHINVLIRIILSYGKMSLIETIGNSOSLENVCRWAYLQOKRPDTHDEYH 354
QY 311 DTAIFLTRODLGATCTDLGMADYGTMCDPKRCSVIEDDGLPSAFTTAHELGHVENMP 370
DB 355 DHAIFLTRDF-GPSGMQ--GVAPVTGMCHPVRSCTLNHEDGFSSAFVAHETGHVLMGE 411
QY 371 HDNV-KVCEEVFGKLRANHMSPTLIQIDRANPWSACSAAITDFLDSGHGDCLLDQP-- 427
DB 412 HDGQGNRCGD---EVLRLGSIAPLVQAAAFHFRHWSRCSQBELSYLHS--YDCLRDDPFT 466
QY 428 -SKPISLPEDLPASYTLSQOCELAFGVSGKPCPTMYQ---CTKLWCTGKAKGOMVCOTR 483
DB 467 HDWP-ALPO-LPGLHYSMNEOCRFDFGLGYMMCTAFRTFPCQKLWCS-HPDNPYFCKTK 523
QY 484 HEPWADGTCGGGKCLKACV-----ERHNLKRVGDSWAKWDPYGPCSRTCCGGV 536
DB 524 KGPDLGTWCAPGKHCFGHCLWLTPDIILK-----DGNWGAWSPPGSCSRTCTGTG 575
QY 537 QLARRQCTNPTPANGKCYEGYRVYKRSCLNLEPCPSSASGKSFREEQCEAFNGY----- 590
DB 576 KPRTRQCDNPHANGRTCSGLAYDFQLCNSQDCPDALA--DFREEQCRQWDLYFEHGA 633
QY 591 -NHSTNRLTAVNVPKYSGVSPRDKCLICRANGTYFYVYAPKYVDGILCS-PDSTSV 648
DB 634 QHH-----WLP-HEHRDAKERCHLYCESKETGEVSVSMKRWVHDGTRCSYKDAFSL 682
QY 649 CVQGGKICAGDGNLGGKRRDFKCGVCGDNKSKKVTGLFTKPMH--GYNFVVAIPAGA 706
DB 683 CYRGDCRKVCGDVGIGSSKQEDKCGCGGDNCHKVYKGTFSRPFKILGIKMEIPIAGA 742
```

```
QY 707 SSIDIRORGYKGLIGDDNYLAKNSO--CKYLLNGHFVYSAVERDLVYVKGSLLYSGTGA 765
DB 743 RHLIQEAD-----TSHHLAVKNLETKGKFLNEENDVDPNKSTFTTAMGVWEYRDE-DG 796
QY 766 VESLQASRPILPLEPLTVLSVYKMTPPRVRYSFYLPKPEPREDKSSHHPKDPGSPVLHNSV 825
DB 797 RETLQTMGPLHGTITVLVIPEG---DARISLTY-----KYMIHEDS 834
QY 826 LSLSNQVQOPDPRPARWVAGSMGPGSCSGSLQKRAVDCRGSGAGORTYPA--CDAAGR 883
DB 835 LNVDDNNVLEDDSVGYEWALKKWSPCSKPGGSGQFTKYGCRRLDHKMVRHRCFCDVSK 894
QY 884 P-----VETQACGEPCPTWELSAWSPCSKSCGR-GFQRRSLKCV-----GHGRLILARDQ 932
DB 895 PKAIRRTCNPQECGQ--PVWVTGEWPCSRSGRTGMQVRSVRCVQPLHNHTTTSVHTKH 952
QY 933 CNLHKBKQELDFCVLRPC 950
DB 953 CN-DARPEGRRACNELC 969
```

Search completed: April 29, 2003, 17:13:55
Job time : 25.7002 secs

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	2482.5	48.1	951	2	T00017	gene ADAMTS-1 prot	
2	1915	37.1	837	2	T00355	hypothetical prote	
3	1604	31.1	550	2	T47158	hypothetical prote	
4	1585	30.7	2165	2	T21371	hypothetical prote	
5	1192	23.1	1205	2	T18517	procollagen N-endo	
6	681	13.2	1558	2	C89114	protein C37C3.6a [
7	681	13.2	2167	2	T34395	hypothetical prote	
8	559.5	10.8	1444	2	T18856	angiogenesis inhib	
9	510.5	9.9	860	2	T16892	hypothetical prote	
10	493	9.5	951	2	T00260	hypothetical prote	
11	450.5	8.7	957	2	T15976	hypothetical prote	
12	411.5	8.0	1059	2	T22545	hypothetical prote	
13	344.5	6.7	571	2	S24789	jararagin C precu	
14	337	6.5	609	2	S55270	cetrocollastatin p	
15	329	6.4	903	2	S60257	metlin alpha - mo	
16	323.5	6.3	617	2	S48160	metalloproteinase	
17	319	6.2	826	2	A0385	monocyte surface a	
18	313	6.1	549	2	S48169	metalloproteinase	
19	308	6.0	789	2	S28259	androgen-regulated	
20	307	5.9	411	1	HY5NFA	fibrolase (EC 3.4.	
21	304	5.9	407	2	S66260	metalloproteinase	
22	300.5	5.8	616	2	A55796	ecarin precursor -	
23	297	5.8	480	1	A30065	trigramin precursor	
24	292	5.7	478	2	JC4880	fibrinolytic prote	
25	290.5	5.6	481	2	JC4342	fibrinolytic prote	
26	289	5.6	414	2	A54609	atrolysin C (EC 3.	
27	286.5	5.5	610	2	JC7530	vascular apoptosis	
28	284	5.5	414	1	HYKSAC	atrolysin C (EC 3.	
29	280	5.4	481	2	S43125	triumucin precursor	

Db **347** GSHTCDTLGMADVGVTCDDPSRSCSVIEDDGLQAAFTTAHELGHVFNMPHDDAKHCASLNG 406
Qy KLRAHHMSPPTLIQTDRANPWSACSAALITDFDSHGDCILLDQPSPKIPISLPEDLPAGASY 442
 :
 :
Db **407** VTGDSLHWASMSLLSDHSQPSWPCSYAWTSTFLDNHGEGICLMDKRPQNPIKPLPSDLPGTYI 466
Qy TLSOQCCELAFGVGSKPCP-YMQYCTKWCKTKACKAQGWQCOTRHFFPWADGTSCSGEKKCLK 501
 :
 :
Db **467** DANOCQQTFEESEKHCPDAASTCTTWCTSTCGLLVCQTKHPFWADGTSCSGEKWCVS 526
Qy GACYERHNHLNKH---RVDSMAWDXPGPCSRCTGGGVLARRCQTNPPTFANGSKYCegy 558
 :
 :
Db **527** GKCVNKTDm-KHEATPVHGSMPGPWGDCSRTCSCGGGYVTMTRECDNPVPKNKGSKYCEGK 585
Qy RVKYRSNCLEPCSSASKGPSFREEOCAFNYNHSTNLRLTLAVAVPKYSVCSPRDKCKL 618
 :
 :
Db **586** RVRTSCNIEDCPN-NKTRFEQCOCAHNEFRKSAPGNPEPTVEWTPTYAGVSPDKCKL 644
Qy ICRANGTGYYFYLVAPKVVDGTLCSPDSTSVCVOGCKIKACGDGNLGSKKRFDKCGVCGGD 678
 :
 :
Db **645** TCEARGIGYFVLQPKVVYDGTPCSPDSTSVCVQCQVKVAGCDRIIDSKKKFKDKCGVCGGN 704
Qy NKSKCKVTGLFTPMGHGVNFVAIPAGASSIIDIRQYKGLIGDDNVYLAKNSOGKLYLN 738
 :
 :
Db **705** GSTCKKMSGIVTSRPGYHDIVTPAGATNIEVHRNQGRSNNGSLAIRAADGTILN 764
Qy GHFVVSVERDLVYKSLLRYSRGTVGESLQASRPPILEPLTVELSVGKMTPPRVRYSF 798
 :
 :
Db **765** GNFTLSLEODLTGKVLRYSGSSAALERIRSFSPLKEPLTIQVLMVGHALRPKIKEY 824
Qy YLPKEPREDEKSSHPRGPPSVLNSVLSLSNQVEQDPQRRPARWAGSWGPCASCSGG 858
 :
 :
Db **825** FMKKKTES-----PNAIPTES-----EWVIEEWGECSKTCGSG. 857
Qy LOKRAVPCRGSAGORTVPACDAAH--RPVTOACE-PCPTWELSANSPOKSCGCRGFQR 915
Db **858** WQRRVQCQRDINGH---PASCAKEVPASTRPCADILFCPHQVQGDWSPCSCTCGKGYKK 914
Qy RSLKCYCHGGHLLARDQCNLHRKPQE-LDFCVLRPC 950
 :
 :
Db **915** RTLKCVSHDGGVLSNESCDPLKKPKHYIDFCTLQCC 950

RESULT 2

T00355 hypothetical protein KIAA0688 - human
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00355
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A>Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: Z14142; MUID:98403880; PMID:9734811
A:Accession: T00355
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-837 <ISH>
A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BAA31663.1; PID:g3327190
A:Experimental source: brain
C:Genetics:
A:Gene: KIAA0688
C:Superfamily: thrombospondin type 1 repeat homology
F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 37.1%; Score 1915; DB 2; Length 837;
Best Local Similarity 46.1%; Pred. No. 1.2e-122;
Matches 381; Conservative 123; Mismatches 267; Indels 52; Gaps 16;

Qy 1 MLILGITLAFAGTAGGEPEREVVVPIRLDPPINERRYWRCPEDSGOGLFIQTAF 60
 :
 :
Db 37 LLLLLALLPSARLASPLPREETIEVPEKLVSVL-----PGSGTFAILLCLRQAQ 88
Qy 61 QEDFYHLTLDAQLFAPAFSTEHLGVLPLOGITGSSDLLRCFSYGDVNABPFSAAYSLC 120

Db 89 GETLLLELQDQSGVQVEGLTVQYLQGAPE-LLGGAPF---GTYLTGTINGDPESVASLHWD 145
QY 121 GG-LRGAFYCGAEVYISPLPNASAPAAQNSOAHLLQRRQVCPGPGSDPTSCGVASG 179
Db 146 GGALLGVQYRGAELHQLPLEGGTPNSA--GGPGAHLRRK-----SPASGGGPNV---196
QY 180 WNPAILRALDPYKPRRAGGESRRRSRAKRFVYIPRYVETLVVADESVMKPHGADLE 239
Db 197 -----KAPLGSPSPRR--RAKRFASLSRFVETLVVADKNAAPFHAGLK 239
QY 240 HYLLTTLTAARLYRHPISILNPINIVVVKLLLRDRSDGPKVTGNAALTIRNFCAMQOKL 299
Db 240 RYLLTVMAAAKAFKHPISIRNPVSLVTVRLVILGSEEGPQVPSAAQTLRSCFAMQRL 299
QY 300 NKVSDKHPIYWDTAFLTFRQDLCAATTCTDLGLMADYVGMCDPKRSCSVIEDDGLPSAFYT 359
Db 300 NTPEDSDPDHFTALFTFRQDLGVSTCDTFLMADYVGMCDPKRSCSVIEDDGLPSAFYT 359
QY 360 AHELGHVFNPHDNVYKVEEVEFGKL-RANHMESPTLIQIDRANPWSACSAIITDFLDG 418
Db 360 AHELGHVFNMLHDNSKPCISLNGPLSTRHVMAVMAHVDPREPWSPCSAAREITDFLDG 419
QY 419 HGDCILLDQSPKISLPEDLPGASVYTLSSQCELAFGVSKPCPYM-QYCTKLWCTGKAKGQ 477
Db 420 YGHCLLDKPEAPLHLVTFPPKQYDADQRCQITFGDPSRHRCQQLPFPCCALMCSGHLNGH 479
QY 478 MVQTRHFPWADGTSCEGKLCILKGACVERHNLNKRVD--GSNAKMDPYGPCSRTC GG 535
Db 480 AMCQTKHSPWADGTPCGPQAQACMGGRCLHMDQLQDPNIPQAGGMPGMPGMDCSRTC GG 539
QY 536 VOLARQCTNPTPANGKGYCEGVKRYKSCNLEPCPSSASGSKSPRECEAFNGYHNSTN 595
Db 540 VQESSDCTRPVPRNGKGYCEGRTRPERSCTEDCP-TGSALTFRREGCAA---YNNRTD 595
QY 596 RLTL---AVAWPKYGSVPDRCKLCIRANGTGYFVYLA PKVVDGTLCPSPDTSVCVQG 652
Db 596 LFKSEFGPMWYPRYTVGAPDQCKLTQCARALGYVYVLEPRVVDGTPCSPDSSVCVQG 655
QY 653 KCTKACGDCNGLSKRRFDKCGVCGDNKCKVTGLTKPHGYNFVVAIPAGASSIDIR 712
Db 656 RCIHAGCDRIIGSKKKFDKCMVCGDGSCKSGSFRKFRGYNNVVTIPAGATHILVR 715
QY 713 QRCYKGLIGDNNYALKNSQKYLNGHFVVSVERDLVVGSL-LRYSGTGTAVSLSQA 771
Db 716 QQNPGF--HRSYTLALKLPDGSVALNGELTLPSPDWWVLPFAVSLRYSGATASETLSG 773
QY 772 SRPLEPTVEVLSVGKMTPRVRYSFYLPKPREDKSHPKD 814
Db 774 HGPLAQPLTVLVAGNQDTRLRYSEFFVPRPTPTPTPTQD 816

RESULT 3
T47158
hypothetical protein DKFzp762c11.0.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47158
R:Blum, H.; Bauearsachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24379
A:Accession: T47158
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-550 <AAA>
A:Cross-references: EMBL:AL162080
C:Experimental source: adult melanoma (Mewo cell line); clone DKFzp762c11.0
C:Genetics:
A:Note: DKFzp762c11.0.1

Query Match 31.1%; Score 1604; DB 2: Length 550;
Best Local Similarity 49.7%; Pred. No. 1.2e-101;
Matches 284; Conservative 106; Mismatches 142; Indels 40; Gaps

RESULT 5
T18517

A;introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 81

Query Match 30.7%; Score 1585; DB 2; Length 2165;
Best Local Similarity 36.7%; Pred. No. 1.2e-99;

Query Match	30.7%;	Score 1585;	DB 2;	Length 2165;
Best Local Similarity	36.7%;	Pred. No. 1.2e-99;		

C:Accession: T18517
R:Collge, A.: Nussgens, B.V.; Lapiere, C.M.
submitted to the EMBL Data Library, February 1996
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A:Reference number: Z18941
A:Accession: T18517
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1205 <COL>
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1
A:Experimental source: skin
C:Genetics:
A:Gene: PC I-NP
C:Function:
A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior to
C:Keywords: hydrolase; metalloproteinase

Query Match 23.1%; Score 1192; DB 2; Length 1205;
Best Local Similarity 31.7%; Pred. No. 3.9e-73;
Matches 329; Conservative 134; Mismatches 403; Indels 172; Gaps 42;

Qy 1 MLLGLILTLAFAGRTAG-----GFEPEREVVVPIRLD----- 32
Db 16 LLLLLLLPADARLAAADDPGGQGHGAERILAVPVRTDAQGLVSHVSAATAPAG 75
Qy 33 -----PDINGRRYYWRGPDGSDGLIFQITAFQEDFYHLHTPDQAFLAPAFST 81
Db 76 VTRRAAPAQIPGLSG-----GSEEDPGGRLEFYNVTVFGRLDLHLRLPNARLVAPGATV 129
Qy 82 EHLGVPLQGLTGSSDLRRCFYSGDV-NAEPDSFAVSLGGLRGAFYGAEBYVLSPL 140
Db 130 EMOGE--SGATRVPEPLGTCLYGVDAEASSVALSNCNGLAGLIRMEEEFFIEPLE 187
Qy 141 NASAPAAQRNSQ--AHLQRRGVP-----GPGSDPTSRGCVASGMNPAIRALDPYKP 193
Db 188 KGL--AAKEAPQGRVHVYHRTPTSRPPPLGGPALDT---GISADSLDSLRAL----- 237
Qy 194 RRAQGESRRRSRRGRKRFVSPTRY-VETILVADENSWKVFHGAAD-LEHVLILLATAAR 251
Db 238 ---GVLEERVNRRRRRHAADDYIEVLLGVDDSVQFQGHTEHVQKYLTLIMNVNE 294
Qy 252 LYRHSILNPINIVVVLRLDRSDGPKV-TGNAALTLNFCAMQKLNKVDKHPYEW 310
Db 295 IYHESLGAHNVVLVRIILLSYKSLSLTEIGNPSOLENVCWVAQVLPKPTDHDYH 354
Qy 311 DTALFTRODLGATTCDTLGMADVTGMDPKRSCSVIEDGLPSATTAHELGHPNMP 370
Db 355 DHAIFLTRQDF-GPSGMQ--GYAPVTGMCHPVRSCITLNHEDGFSSAFVAHETGHLGME 411
Qy 371 HDNV-KVCEEVFGKLRANHMSPTLIQIDRANPHSACSAALITDFLDSGHGDCLLDOP-- 427
Db 412 HDGQGNRCGD---EVLRLGSIAPLVQAFAHFRHWSRCSQOELSRYLHS--YDCLRDDPFT 466
Qy 428 -SKPISLPEDLPGLASYTLISOCELAFVGSKPCPYMOY---CTKLWCTGKAKGMVQOTR 483
Db 467 HDWP-ALPQ-LPLGLHYSNNEQCFDEGLGYMCTAFRTFDEPKQLWCS-HPDNPFYCKTK 523
Qy 484 HFPWADGTCGEGKLCCLKACV-----ERHNLNKHVRDGSWAKWDYGPSCSTCGGV 536
Db 524 KGPPLDGTMCAPGRKCFKGHCIWLTPDLTKR-----DCNMGAWSPFGSCSRTCTGV 575
Qy 537 QLARQCTNPTPANGKCYGVRVYKSCNLEPCPSSASGSKSFREEOCEAFNGY----- 590
Db 576 KFRTRQCDNPANGRTCSGLAYDFQLCNSQDCPDALA--DFREOQCRMDLYFEGHDA 633
Qy 591 -NHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYFYVLAPKVDGTLCSS-PDSTSV 648
Db 634 QHH-----WLP-HEHRDAKERCHLYCESKETGEVVSCKRMVHDGTRCSYKDAFSL 682
Qy 649 CVQGCKIKAGCDNGLSKKRFDKGCVCGGDNKSCKKVTGLFTKPMH--GYNFVVAIPAGA 706
Db 683 CVRGDCRVGCDVIGSSKQEDKCGVCGGDNSHCKVWVAGTFSRSPKILGYIKMFEIPAGA 742

Qy 707 SSIDIRQGYKGLIGDDNYIALKNSQ-GKYLNGHFVYSAVERDLVVKGSLLRYSGTGTA 765
Db 743 RHLLIQEAD-----TSHLLAVKNEGTGKFLLENDVDNPKTFIAMGVWEYERDE-DG 796
Qy 766 VESLQASRPILPTVEVLSVGRKMTPPRVRYSYFLPREDKSSHPKDPGRGPSVLHNSV 825
Db 797 RETIQTMGPLHGTITVLVPIEG---DARISLTY-----KYMIIHEDS 834
Qy 826 LSLSNQVQPDPRPPARVWAGSWGSPCSASCSGLQKRAVDCRGSAGQRTVPA--CDAHR 883
Db 835 LNVDDNNVLEDSVGYEWALKKWSKPCSKCGGSGQFTKYGCCRRLDHKMHVRGFCDSVSK 894
Qy 884 P-----VETQACGEPCTWELSAWSPCSKSGR-GFORRLKCV-----GHGGELLARDQ 932
Db 895 PRAIRTCNPQECSS--PWWTGWEPSCSGRTGMQVSRVCVQPLHNTTTSVHTKH 952
Qy 933 CNLHRKPOELDFCVLRPC 950
Db 953 CN-DARPEGRACNRELC 969

RESULT 6
C89114
protein C37C3.6a [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C89114
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: C89114
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1558 <STO>
A:Cross-references: GB:chr_V; PIDN:AAC25867.1; PID:g3294501; GSPDB:GN00023; CBSP:C37C
C:Genetics:
A:Gene: C37C.6a
A:Map position: 5

Query Match 13.2%; Score 681; DB 2; Length 1558;
Best Local Similarity 33.7%; Pred. No. 3.8e-38;
Matches 162; Conservative 63; Mismatches 176; Indels 80; Gaps 19;

Qy 517 GSWAKWDPYCPGSRCTCGGVQVLARQCTNPTPANGKYCEGVYKVRKSCNLEPCPSSASG 576
Db 77 GNMGPVWVPENECSRCSGGGVQLEKQCSGD-----CTGASVRIISCNLNAC---ESG 125
Qy 577 KSPREQCEAFN-----GYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYFYVL 631
Db 126 TDFRACQCKFNDALDGNVHK-----WTP-YKG---KNKELCKPESGNFYKWK 172
Qy 632 APKVVDGTLCPDSTSVQVQKCIKAGCDNGLSKRPFKDKGCVCGGDNKSCKKVTGLFTK 691
Db 173 ADKVVDGTRKDSKNDICVDGECPLVGCDCGLGSSLFKDKCGCDGSGTCTIEGREDE 232
Qy 692 P--MHGYNFVVAIPACASSIDIRQYKGLIGDDNYIALKNSQKYLNGHFWVSAVERD 749
Db 233 RNLSPGYHDIILPEGATNIKIQEAR-----KSTNNLALKNSGDHLYLNGSLIO-VEKE 286
Qy 750 LVVKGSLRYSGTGTAVERSLQASRPILPTVEVLSVGRKMTPPRVRYSYFLPREDE-- 807
Db 287 VEVGGTIFVYD--DAPEFELSAQGPLSELTVALLFKGRSDTAIKYFISIPLEERVDM 344
Qy 808 -----KSSHPKDPGRPSVLHNSVLSLNOVEQDDRP-----PAR 842
Db 345 YKEDNTPCPSVSCGKGVQTRNLVYCIDGKNKGRVEDDLCEENNATKPEFEKSETVDCEAE 404
Qy 843 WVAGSWGPCASCG-SGLQKRAVDC-RGSAGQRTVPAQD---AAHRPVETOACGE-PCPT 896
Db 405 WFTGDMESCSSTCGDQGGQYRVVYCHQVANGRRVTVEDGNCTVPRPPVQTCNRFACPE 464

QY 517 GSWAKWDPYGPCSRTCCGGVQLARQC-----TNPTPANGKYCEGVRVYKRSCHLEPCPS 572
DQ 51 GWTWKTAF---SRCCGGGVTSQERHCLQORRKSVPNGNRTCTGTSKRYQLCRVQECF- 106
QY 573 SASGKSFREOCEAFNG--YHNSTNRLTLAVAWVPKYSGVSPRD-----KCLICRAN 623
DQ 107 -PDGRSFREOQCVSNHYNRTHQ-----WKPLY-----PDDVHLSRCPDLHC-TT 154
QY 624 GTCYFYVLAPKVVDTGLCS-PDSTSVCVQKCIKAGCDGNLGSKKRFDKCGVCGDNKSC 682
DQ 155 VDGQQLMVP-ARDGTSKLDLRCVGVSGRCEPTGCDGVLFSTHTLDKCGICQGDGSSC 213
QY 683 KVTGLFTK-PMH-CYNFWFAIPAGASSIDIRQCYKGLIGDNDVNLKNSOGKYLNGH 740
DQ 214 THVTGNRYKGAHLGYSLVTHIPAGARDIQIYERKKA-----DVLALADAGYFFNGN 268
QY 741 FVSAVERDLVVKGLLSYSGT-----GTAVESLQASRPILEPLTVEVLSV-GKMTPPRVR 795
DQ 269 KYVDS-PKNFNIAGTVVYKRRPMDVYETGIEYIQAQGTNOGLNVMVWQNGK--SPSIT 325
QY 796 YSYLPKPEPRDKS----- 809
DQ 326 FEYTLQPHESRPOPIYGFSESAESQGLDGLMGPFIHNGSLYQOASERLGLDNL 385
QY 810 -SHP----- 812
DQ 386 FQHPGLDMELGPSOQETNEVCEQAGGACGPPRGKGRDRNVTGPTLTKDKDEEVD 445
QY 813 -----KD-----PR----- 816
DQ 446 HFASQEFSSANISDOLLGAGSLDKFTLNETVNSIFAQAPRSSLAESFFVDYNEGA 505
QY 817 GPSVLHNSVLSL-----NOVEQPDPRPPA----- 841
DQ 506 GPYLLNGSYLELSSDRVANSSEAPPNVSTSLTSAGNRTHKARTPKARKQGVSPADM 565
QY 842 ----- 842
DQ 566 YRWKLSHSPCSATCTTGVMASAYAMCVRYDGVVEVDSYCDALTRPEPVHFCAGRCQPR 625
QY 843 WVAGSWGPCSASCSSGLQKRAV----- 864
DQ 626 WETSSWSCSRTCCGGYQFVVRVWKMLSPGDFSSVSDICEAAEAVRPEREKTCTRNPA 685
QY 865 -----DCRGSAGORTV-----PACDAARHPVETQAC-GEPCP-TWELSA 901
DQ 686 GPQWEMSEWSECTAKGERSVVTDRIRCSDEKLCDDNTRPVGEKNCTGPPCDRQWTVSD 745
QY 902 WSPCKSGGGRGFRSLKCVCHGGRLLARDQCNLHRYPQELDFCVLRPC 950
DQ 746 WGPCSGCGGORTIRHYVCKTSDGRVVPESQCMETKPLAIHPCGDKNC 794

RESULT 11

Ti5976

hypothetical protein F08C6.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: Ti5976

R:Bentley, D.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid F08C6.

A:Reference number: Z18440

A:Accession: Ti5976

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-957 <BEN>

A:Cross-references: EMBL:U29378; NID:9868184; PID:9868185; PIDN:AAA68721.1; CESP:F08C6.1

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F08C6.1

A:Introns: 23/1; 135/3; 220/3; 285/3; 325/2; 376/3; 461/2; 498/3; 532/3; 627/2; 714/1; 7

Query Match 8.7%; Score 450.5; DB 2; Length 957;
Best Local Similarity 22.7%; Pred. No. 1.1e-22;
Matches 227; Conservative 106; Mismatches 328; Indels 339; Gaps 47;
QY 57 ITAFQEDFYLHLPDAQFLAPAFSTHGLVPLGLTGGSSD-----LRRFYSGDVNAEPD 112
DQ 61 IQAFNKKYINLSL---EXTLAKLLSSGVTVVKKNEKGGSLDFTGSLDSCYH---HYGEK 114
QY 113 SFAVSLCCGGL-----RGAFGYRGAEEYVISLPNASAPAAOANSQ-GAHLQORG 161
DQ 115 VYAAISNCDAIRIYESNRKGGKGTVDGELIVVHFPPDHAHRSKRATENGAAHVYKRE 174
QY 162 VPGGSGDPTSRGCVAGWNPAILRALDPYKRRAGFESRRSRGRKRFRVSIPIR--- 218
DQ 175 T---LAGEPKDFGGLDN-----VVTESLVEDESALFEDV-----FVTGQRLTQ 215
QY 219 ---YVETLVVADESVMKF---HG--AD--LEHYLLTLTAARLYRHPSTILNINIVY 266
DQ 216 QSLIVLAVFVDENLWRHFSKHHGMADKKLDYTLTLNINQIYQTPASPPLTFRV 275
QY 267 VKVLLLRDRDSG---PKVTGNAALTNRFCAMQKKLNKVSOKHPEYWDTAILFTQDL- 321
DQ 276 IRYEVLTRPSALAGYLHNHNAQMYLDRFCRYORN-AYRD-----WDHAILMTGYDIH 329
QY 322 CGATTCTILGMADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKKVCEVF 381
DQ 330 RGAGRSISGIALDGMCDPWNCTLAEGLDFTSAGTIGHELGHRTFLKDH----- 381
QY 382 GKLRANHMSPTLIQIDRANPWSACSAIITDPLDSHGDCLLDQPSKPISLPEDLPAS 441
DQ 382 -----KSDTLF----- 387
QY 442 YTLSSQCELAFGVGSKPCPYMQYCTKMTCTKRAKGOMVCQTRHFPWADGTSCGEKLCUK 501
DQ 388 -----TGTFCGSKWCOL 400
QY 502 GACVERHNLNK-----HRVDGSAKWDYPGCSR-TCGG-----GVQLA 539
DQ 401 GRCPWPTGNETQPTVOHVAVPVVITLPSRIDGSWGGA-TICSOCTCNGILGSLVGLA 459
QY 540 RROCTNPTPANGKYCEGVRVYKRSCHLEPCPSASGSKFREEQC-----EAFNGY 590
DQ 460 RRTCSAPYPANGSGDCVGTSTRAVLCSRCGRASKSVDEYISDKCMEQKRLKNDRELTKG 519
QY 591 NHSTNRLTLAVAWVPKYSGVSPRDKCKLICRAN---GTGYFY-VLAPKVVDTGLCSPDST 646
DQ 520 GSQNLNF-----PORACKVFCVQOYHGSQNRNYRFFGDNLPDGTSCGYD-- 563
QY 647 SVCVOGCKIKAGCDGNL-----GSKKRFDC-GVCGGD 678
DQ 564 RYCLDGECLALNCNNALLSRDQSCPTDCTPITDOSSSVYRQGWGTSLWTSCTATCGG 623
QY 679 ----NKSKKVTGLTKPMHGYNFVVAIPAGASSIDIRQGYKGLIGDNDVNLKNSOGK 734
DQ 624 YRKRNRAC-SITG-----QCEGNEDETEVCSSCPSV-----LRVGNWST 664
QY 735 YLLNGHFVYSAVERDLVWKVGSLLRYSGTGTAVESLQASRPILEP---LTVEVLSVGKWT 791
DQ 665 WTEWNNHCVSC-----GRGSOARY-----RKCLSPHRTLAFDPCGNGKVT- 704
QY 792 PVRYSFYLPKPEPRDKSHKPRGPSVYLHNSVLSLSNQVEQPDPRPP-----ARWA 845
DQ 705 NELRITFF-----KARSYIMCSVRCNKIKRTISEKNIEVRSCDNGPCNAIGVWGTW- 756
QY 846 GSWGPCSACSGSG--LQKRAVD---CRGSAGORTVPACDAARHPVETQACPEPT---W 897
DQ 757 GWSWTSCTSGGTGLVTRORTNREPCDSAHERR--SCNVA-----TCQNDGIW 803
QY 898 EL-SAWSPCSCGGRGFORRLSKVCHG--GRLLARDQCN 934
DQ 804 SLWNWSDCSRVCVCGGLRSRSCFCGSGCMGASSEQFCN 843

RESULT 12

T22545

hypothetical protein F53B6.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22545

R:White, S.

Submitted to the EMBL Data Library, October 1996

A:Reference number: Z19578

A:Accession: T22545

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1059 <WIL>

A:Cross-references: EMBL:Z81086; PIDN:CA03121.1; GSPDB:GN00019; CESP:F53B6.2

A:Experimental source: Clone F53B6

C:Genetics:

A:Gene: CESP:F53B6.2

A:Map position: 1

A:Introns: 38/3; 92/3; 131/3; 169/1; 236/2; 282/3; 349/2; 376/2; 420/1; 576/3; 759/3; 84

Query Match

Best Local Similarity 8.0%; Score 411.5; DB 2; Length 1059;

Matches 112; Conservative 53; Mismatches 128; Indels 143; Gaps 16;

QY 518 SWAKWDYPCSRCTCGGQVQLARRQCTNPTPANGKCYCEGVVRVYKSCNLEPCPSSASGK 577

Db 25 SWAWSPSSCTKTCGGVSRQLRCLT-----SKCSGESVRFKYCAQKTCESK--R 75

QY 578 SFREEQCAFNGYNHSTNRLTLAVAVPKYSGVSPDKCLICRANGTGYFYVLAPKVD 637

Db 76 LARDTICGG-----EEIVSRGQCEVVCVRSRLTGANFLW--RVDD 112

QY 638 GTLC-SPDSTVCYQGIKAGCDNLGSKRFPDKCGVCGDNKSKVKVTLFTKPMHY 696

Db 113 GTPCAATSRVACSGSQIYVCCDGLTSSFRPDACVCGGRGDT----- 158

QY 697 NFVVAIPAGASSIDIRQRYKGLIGDNYLALKNSQKYL--LNGHFWVSAYERDLVVK 754

Db 159 -----DN-----GKFIWKVSEYTCASNCDDIVD- 183

QY 755 SLLRYSGTGAVESLOASRPILPTVEVLVS-GKMTPPRVRYFYLPKEPRDKSSH 813

Db 184 ---WSGAGRSTASTS-----QPIVVCVNAITGRVVPKEKLCADKLRP----- 222

QY 814 DPGSVLHNSVLSLSNQVEQDDRP-----PARWAGSMGPGSCSGGLQKRAYDCR 867

Db 223 -----VEARPCMLICPSRWMAADWTECVPHCGEGTRKREVVCV 261

QY 868 GSAGORTVPA---CDAARPVETQAC-GEPCPTWELSAWSPCKSCGSGRFPORSLKCVG 922

Db 262 QTAHNVTVHPDTCENGTRPAAEENCYSTSCGRWEAGKSKCTATSCGQGVRRRHVACVG 321

QY 923 -----HGGRLLRADC 933

Db 322 GSDCDEGRPRQETTC 337

RESULT 13

S24789

jararagin C precursor - jararaca (fragment)

N:Alternate names: single chain botrocetin

N:Contains: disintegrin-like 28k protein; hemorrhagic proteinase (EC 3.4.24.-)

C:Species: Bothrops jararaca (jararaca)

C:Date: 20-Feb-1995 #sequence_revision 29-Aug-1997 #text_change 09-Jun-2000

C:Accession: S24789; JC2245; A4463; A37958; JC2373

R:Palme, M.J.I.

Submitted to the EMBL Data Library, August 1992

A:Reference number: S24789

A:Accession: S24789

A:Molecule type: mRNA

A:Residues: 1-571 <PAT>

A:Cross-references: EMBL:X68251; NID:G62467; PID:G62468

R:Usami, Y.; Fujimura, Y.; Miura, S.; Shima, H.; Yoshida, E.; Yoshioka, A.; Hirano, K
Biochem. Biophys. Res. Commun. 201, 331-339, 1994

A:Title: A 28 kDa-protein with disintegrin-like structure (jararagin-C) purified fro

A:Reference number: JC2245; MUID:94256999; PMID:8198592

A:Accession: JC2245

A:Molecule type: protein

A:Residues: 360-571 <USA>

A:Experimental source: venom

R:Palme, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.

J. Biol. Chem. 267, 22869-22876, 1992

A:Title: Purification, cloning, and molecular characterization of a high molecular we

ily.

A:Reference number: A4463; MUID:93054601; PMID:1385408

A:Accession: A4463

A:Molecule type: mRNA

A:Residues: 1-23,'Q',25-92,'G',94-131,'G',133-169,'Q',171-571 <PA2>

A:Cross-references: GB:X68251

A:Experimental source: venom gland

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIP:118104)

R:Fujimura, Y.; Titani, K.; Usami, Y.; Suzuki, M.; Oyama, R.; Matsui, T.; Fukui, H.;

Biochemistry 30, 1957-1964, 1991

A:Title: Isolation and chemical characterization of two structurally and functionally

A:Reference number: A37958; MUID:91129280; PMID:1993206

A:Accession: A37958

A:Molecule type: protein

A:Residues: 360-372,'E',374-378,'X',380-384 <FUJ>

A:Note: 361-Val was also found

C:Comment: Inhibits collagen- and ADP-induced platelet aggregation.

C:Superfamily: mouse meltrin alpha; disintegrin homology

C:Keywords: hydrolase; metalloproteinase; venom; zinc

F:360-571/Product: jararagin C #status experimental <MAT>

F:362-444/Domain: disintegrin homology <DIS>

F:295,299,305/Binding site: zinc (His) #status predicted

F:296/Active site: Glu #status predicted

Query Match

Best Local Similarity 6.7%; Score 344.5; DB 2; Length 571;

Matches 160; Conservative 88; Mismatches 256; Indels 201; Gaps 35;

QY 12 AGTAGGFEPREVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAFQEDF-YLHLTP 70

Db 1 ATRPKGAVPKYEDA--MQEYKNGCEVPVLHLEKNG-----LFSKDYSEIHYSP 49

QY 71 DAQFLAPARSTELHGLVPLQGLGSSDLRCFYSGDVNAEPDPSFAVSLCGGLRGAFGYR 130

Db 50 DGREITTYPPVED-----HCYVHGRIENDADSTASISACNLKGLFKLQ 93

QY 131 GAQYVISP--LPNASAPAAQRNSQGAHLQRRGVPGGSPGSDPTSRGCVASGWN----- 181

Db 94 RETYFIEPLKLPDSEAH-----VFKYENVE--KEDEAPKMGVTONKKSVEPIKK 142

QY 182 -----FAILRALDPYKPRRAGFGESRRSRRAKRFVSIPIRYVETLVVADESVMKPHG 235

Db 143 ASQLAFTAEQORYDPYK-----YIEFFVVVDGTVTKNN 176

QY 236 ADLEHYLLTLATAARLYRHPSILNPI-NIVVVKVLL--LRDRSGPKVT--GNAALTLR 290

Db 177 GDLD-----KIKARMYELANIVNEIFRYLYMHVALGLEIWSNGDKITVRPDVDTLN 229

QY 291 NFCAWQKLNKSKDKHPEYWDTAIFTRQDLCGATTCTDLGMADVGTMTCDPKRSCSVIED 350

Db 230 SFAENRKTDLTLRKHH---DNAQLLTAIDFNG---PTIGVAYIGSMCHPKRSVGIVQD 281

QY 351 ---DGLPSATTAHGLHGFVNMFDNVKVEEFGKLRANH---MMSPTLIQIDRANPWS 404

Db 282 YSPINLVVAVIMAHGMHNLGIHHT-----GSCSGDYPICMGPT-INSNEPSKFFS 332

QY 405 ACSAAIITDFLDGSHGDCILDDQP-----SKPISLPEDLPGASVYTLSSQCELAFGVGRSP 458

Db 333 NCSYIQWDFIMNHNEPCEINELGTDIISPPVCGNELL-----EVGECD----- 378

QY 459 CPYMQYCTKLWCTGKAKGQMVCTQTRHFPFWADGTSCGEGKLCCLKGACVERHNLNKHVD-- 516

Db 379 CGTPENQNECCDA-----ATCKLK-----SGSQCGHD-----CCEQCKFSKSTECR 422
QY 517 GSAKWDYPCSRCTCGGVQVLARROCTNPTPANGKYGCEGVRYKRSNLEPCP-----571
Db 423 ASMECDPAHCHGQS-----ECPADVPHKNGQPCLD---NYGYNCNGCPTMYHQ 471
QY 572 -----SSASGKSFRECEAFNGYNIHSTNRLTLAVAVPKYSGVSPRD-KC-KLIC 620
Db 472 CYALFGADVYAEEDSCFDKDNKNGYGYCKRKGKIPCA-----PEDVKGRLYC 522
QY 621 RANGTG-----YFYVLAPK-----VVDGTLCSPDSTSVCVQKCI 655
Db 523 KDNSPGQNNPKRMFYNDDEHKGMVLPGTKCA--DGKVCNSGHC 565
RESULT 14
S53270
catrocollastatin precursor - western diamondback rattlesnake
C:Species: Crotalus atrox (western diamondback rattlesnake)
C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C:Accession: S55270
R:Zhou, Q.; Smith, J.B.; Grossman, M.H.
Biochem. J. 307, 411-417, 1995
A>Title: Molecular cloning and expression of catrocollastatin, a snake-venom protein from
A:Reference number: S55264; MUID:95251603; PMID:7733877
A:Superfamily: mouse meltrin alpha; disintegrin homology
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-609 <ZHO>
A:Cross-references: GB:U21003; NID:9710353; PIDN:AAC59672.1; PID:g710354
C:Superfamily: mouse meltrin alpha; disintegrin homology
Query Match 6.5%; Score 337; DB 2; Length 609;
Best Local Similarity 22.8%; Pred. No. 3.5e-15;
Matches 150; Conservative 90; Mismatches 259; Indels 160; Gaps 30;
QY 50 DQGLIQTAFQEDFYLLHTPDQAFLAPAFSTEHLGVPLQ-----GLTGGSSDLRRFCYSGDNA 109
Db 52 EDAMQYELKVNQEPVYVHLGKNGKLFSDYSETHYSPDGREITVPLVEDHCYHGRLEN 111
QY 110 EPDSEAAVSLCGGLRGAFYGAEEYVISP--LPNASAPAAQNSOGAHLQRRGVPGGPS 167
Db 112 DADSTASISACNLGKHFKLOGEMLLIEPLKLPDSEAHVAYKYE-----KE 160
QY 168 GPTSRGCVASGNPAILRALDYPKPRAGFGESRRSRRAKRFVSP-----RY 219
Db 161 DEALKMCGVTQNW-----ESYEPIK-----KASQIVTAHQKYNPFRF 199
QY 220 VETLVVADESVMKFGADLEHY---LLTLATAARLYRHPSILNPINIVVVKLLRRDRD 276
Db 200 VELFLVVDKAMVTKNNGDLDKIKTRMYEIVNTVNEYRYM---IHVALVGLIWSNED 255
QY 277 SGPKVT--GNAALTNRNCAWOKLNKYSKHPYWDYTAFLTRODLGATTCDTGLHAD 334
Db 256 ---KIIVKEPAGVTNAGFEWARKTDLITRKKH-----DNAQLTAIDL-----DRVIGLAY 303
QY 335 VETMCDPKRSCSVIED---DGLPSAFTTAHELGHVFNPHDN-VKCEEVFGKLRANHMM 390
Db 304 VGSMDPKRSTGIQDYSEINLVAVIMAHENGHNINHDGSCGCGDYACIMPEISP 363
QY 391 SPTLIQIDRANPWSACSAAITDGLDSHGDCLLDOP-----SKPISLPEDLPQASTYL 444
Db 364 EPSTF-----FNSCSYFECWDFIMNHNPECILNEPLGTDIISPPVCGNELL-----EV 411
QY 445 SQOCELAFGVGSKPCYMOYCTKLTCTGKAKQMWQCTPHFPWADGTCGEGKCLKKAC 504
Db 412 GECDD-----CCTPENCNECCDA-----ATCKLK-----SGSQCGHD-----C 446
QY 505 VERHNLNKRHVD--GSWAKWDYPCSRCTCGGVQVLARROCTNPTPANGKYGCEGVRYKY 562
Db 447 CEQCKFSKSTECRASMSCEDPAEHCQTGSS-----ECPADVPHKNGQPCLD---NY 495

QY 563 RSNLEPCP-----SSASGKSFRECEAFNGYNIHSTNRLTLAVAVPKYS 608
Db 496 GYCYNGNCPIIMYHQCYYDLFGADVYAEEDSCFERNOKNGYGYCKRKGKIPCA-----549
QY 609 GVSPRD-KC-KLICRANGTG-----YFYVLAPK-----VVDGTLCSPDSTSVCVQKCI 655
Db 550 ---PEDVKGRLYCKDNPSGQNNPKRMFYNDDEHKGMVLPGTKCA--DGKVCNSGHC 603
RESULT 15
S60257
meltrin alpha - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
C:Accession: S60257
R:Yagami-Hiromasa, T.; Sato, T.; Kurisaki, T.; Kamiyo, K.; Nabeshima, Y.I.; Fujisawa
Nature 377, 652-656, 1995
A>Title: A metalloprotease-disintegrin participating in myoblast fusion.
A:Reference number: S60257; MUID:96026308; PMID:7566181
A:Accession: S60257
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-903 <YAG>
A:Cross-references: EMBL:D50411; NID:g1054586; PIDN:BAA08912.1; PID:g1054587
C:Superfamily: mouse meltrin alpha; disintegrin homology
F:421-503/Domain: disintegrin homology <DIS>
F:349/Active site: Glu #status predicted
Query Match 6.4%; Score 329; DB 2; Length 903;
Best Local Similarity 21.5%; Pred. No. 2e-14;
Matches 159; Conservative 74; Mismatches 245; Indels 262; Gaps 27;
QY 2 LLIGLTLTAPAGRTA-----GGPEPEREVVPIRLDPIINGRRYVWRGPDSDGQG 52
Db 15 LLLALAGALLAPRAAGMSLWDQGAIVEARASL--LSKDPGIPQSI-----PAKHDPDV 68
QY 53 LIFOITAFQEDFYLLHTPDQAFLAPAFSTEHLGVPLQ-----GLTGGSSDLRRFCYSGDV 107
Db 69 LTVQLQLESRDLLSLERNEGLANGFTETHY---LQDGTDVSLTRNHTD--HCYIYHGV 123
QY 108 NAEPSFAAVSLCGGLRGAFYGAEEYVISP--LPNAS-----APA-AORNQOGAHLQRRG 161
Db 124 QGDAASVYSLSTCSDLRGLIMFNKTYSLPEMKNKNTSYKLVPAESMTNIOGL-----176
QY 162 VPGGPGDPTSRGCVASGNPAILRALDYPKPRAGFGESRRSRRAKRFVSIPIYVE 221
Db 177 -----CSQHNKSNLTWEDVSP-----GTSQMRARHK---RETLMKTIYE 215
QY 222 TLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVVKLLRRDRDSGPV 281
Db 216 LVIVADNREFQRCQKLEKVKQRLIEIANHVDFKFRPLN-IRIVLVGVEVWNDIDK--CSI 273
QY 282 TGNAAITLRFNCAWOKLNKYSKHPYWDYTAFLTRODLGATTCDTGLMADVGTWCDP 341
Db 274 SQDPFTRLHEFLDWRKIKLLPRKSH---DNAQLISGYFQGT---TIGMAPIMSMCTA 325
QY 342 KRSCSVI---EDDGLPSAFTTAHELGHVFNPHDN-VKCEEVFGKLRANHMSPTLIQI 397
Db 326 EQSGGVVMDHSDSPGAAVTLAHELGHVFNPHDNVTLERGCSCRMAGKGCIMNPS-----381
QY 398 DRANP-----WSACSAAITDGLDSHGDCLLDOP-----SKPISLPEDLPQAS---441
Db 382 -TGFPFPMVTFSSCRKDLASLEKGMCMCLNLPVQAFGRKCGNGYVEEGECDCGE 440
QY 428 -----YTLSSQOCELAFGVGSKPCYMOY-----SKPISLPEDLPQAS---441
Db 441 PEECTNRCNATTCTLKPDVCAHQCCEDCQLKPPGTACRGSNSCDLPEFCTGAPHC 500
QY 442 -----YTLSSQOCELAFGVGSKPCYMOY-----SKPISLPEDLPQAS---464
Db 501 PANYLHDGHPCCQVDCYCYNGICQTHEQOCQVTLWGPAGKAPGAGICFERVNSAGDYGNC 560
QY 465 -----CTKLWCTGKAK-----GOMVQTRHFPW 487

Db 561 GKDSKSAFAKELRDAKCGKIQCQGGASRPVIGTNAVSIETNIPQOEGGRILCRGTHVYL 620
Qy 488 AD-----GTSCEGKCLKAGACVERHNLNKHVRVDSWAKWDPYGPCSRTC--GGV 536
Db 621 GDDMPDPLVLGATKCAEGKICLNRRQCNTISVFGVHK-----CANQCHGRGV 667
Qy 537 QLARROCTNPTPANGGKYCE 556
Db 668 CNRRKNC-----HCE 677

Search completed: April 29, 2003, 17:13:17
Job time : 44.9736 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:22:38 ; Search time 26 Seconds
(without alignments)
2927.820 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 950
Sequence: 1 MLLIGILTLAFAGTAGGFE.....DQCNLHRKQELDFCVLRPC 950

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 301932 seqs, 80129803 residues

Word size : 0

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	931	98.0	950	10 US-09-965-631-4	Sequence 4, Appli
2	408	42.9	823	9 US-10-163-316-2	Sequence 2, Appli
3	345	36.3	367	10 US-09-965-631-6	Sequence 6, Appli
4	300	31.6	321	10 US-09-965-631-2	Sequence 2, Appli
5	17	1.8	608	10 US-09-803-589-2	Sequence 2, Appli
6	17	1.8	608	10 US-09-803-589-8	Sequence 8, Appli
7	17	1.8	608	12 US-10-105-929-13	Sequence 13, Appli
8	17	1.8	727	9 US-10-097-597-1	Sequence 1, Appli
9	17	1.8	727	9 US-10-097-597-12	Sequence 12, Appli
10	17	1.8	727	9 US-10-097-580-1	Sequence 1, Appli
11	17	1.8	727	9 US-10-097-580-12	Sequence 12, Appli
12	17	1.8	727	10 US-09-445-023A-1	Sequence 1, Appli
13	17	1.8	727	10 US-09-445-023A-12	Sequence 12, Appli
14	17	1.8	950	10 US-09-321-987B-4	Sequence 4, Appli
15	17	1.8	967	12 US-10-105-929-2	Sequence 2, Appli
16	17	1.8	968	9 US-10-163-316-7	Sequence 7, Appli
17	14	1.5	518	10 US-09-803-589-10	Sequence 10, Appli
18	14	1.5	551	12 US-10-105-929-16	Sequence 16, Appli
19	13	1.4	837	9 US-10-174-590-352	Sequence 352, App

20	13	1.4	837	9	US-10-176-758-352	Sequence 352, App
21	13	1.4	837	9	US-10-175-737-352	Sequence 352, App
22	13	1.4	837	9	US-10-173-706-352	Sequence 352, App
23	13	1.4	837	9	US-10-175-738-352	Sequence 352, App
24	13	1.4	837	9	US-10-175-752-352	Sequence 352, App
25	13	1.4	837	9	US-10-176-482-352	Sequence 352, App
26	13	1.4	837	9	US-10-176-757-352	Sequence 352, App
27	13	1.4	837	9	US-10-176-913-352	Sequence 352, App
28	13	1.4	837	9	US-10-180-552-352	Sequence 352, App
29	13	1.4	837	9	US-10-180-557-352	Sequence 352, App
30	13	1.4	837	9	US-10-173-700-352	Sequence 352, App
31	13	1.4	837	9	US-10-174-572-352	Sequence 352, App
32	13	1.4	837	9	US-10-174-579-352	Sequence 352, App
33	13	1.4	837	9	US-10-174-582-352	Sequence 352, App
34	13	1.4	837	9	US-10-174-588-352	Sequence 352, App
35	13	1.4	837	9	US-10-175-739-352	Sequence 352, App
36	13	1.4	837	9	US-10-175-740-352	Sequence 352, App
37	13	1.4	837	9	US-10-175-743-352	Sequence 352, App
38	13	1.4	837	9	US-10-176-488-352	Sequence 352, App
39	13	1.4	837	9	US-10-176-492-352	Sequence 352, App
40	13	1.4	837	9	US-10-176-747-352	Sequence 352, App
41	13	1.4	837	9	US-10-176-750-352	Sequence 352, App
42	13	1.4	837	9	US-10-176-985-352	Sequence 352, App
43	13	1.4	837	9	US-10-176-987-352	Sequence 352, App
44	13	1.4	837	9	US-10-176-991-352	Sequence 352, App
45	13	1.4	837	9	US-10-176-992-352	Sequence 352, App

ALIGNMENTS

RESULT 1
US-09-965-631-4
; Sequence 4, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-4

Query Match	98.0%;	Score 931;	DB 10;	Length 950;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 931;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	20	EPERVVVPIRLDPDINGRRYWRGPGSDGGLFQITAFQEDFYHLHTPDQFLAPAF 79		
Db	20	EPERVVVPIRLDPDINGRRYWRGPGSDGGLFQITAFQEDFYHLHTPDQFLAPAF 79		
QY	80	STHLGVPLQGLTGSSDLRCFYSGDVNAEPDFAAVSLCGGLRGAFYRGAEVVISPL 139		
Db	80	STHLGVPLQGLTGSSDLRCFYSGDVNAEPDFAAVSLCGGLRGAFYRGAEVVISPL 139		
QY	140	PNASAPAAQRNSQGAHLQRGVPGGSDPTSCRGVAGSNWNPAILRALDYPKPRRAGFG 199		
Db	140	PNASAPAAQRNSQGAHLQRGVPGGSDPTSCRGVAGSNWNPAILRALDYPKPRRAGFG 199		
QY	200	ESRRRRSGRAKRFVSIPIRYVETLVVADESVMVKFHGADLEHYLLTATLAARLYRHPISIL 259		
Db	200	ESRRRRSGRAKRFVSIPIRYVETLVVADESVMVKFHGADLEHYLLTATLAARLYRHPISIL 259		
QY	260	NPINIVVYKVLRLDRDSDGPKVTGNAALTLRNFCAWQKLNKVDKHPYWDTAILFTQ 319		

Db	260	NPINIVVVKVLLRRDSGPKYGTNAALTNRFCWAQOKLNKYSKHPYMDTALITFRQ	319
Qy	320	DLCGATTCDTGLMADVGTMCDPKRCSVIEDDGLSPSAFTTAHELGHVFNMPHDNVKVCCE	379
Db	320	DLCGATTCDTGLMADVGTMCDPKRCSVIEDDGLSPSAFTTAHELGHVFNMPHDNVKVCCE	379
Qy	380	VFGKLRANHMWSTLTLOIDRANPWSACSAALITDFLDSGHGCDLDOQSKPISLPEDLPG	439
Db	380	VFGKLRANHMWSTLTLOIDRANPWSACSAALITDFLDSGHGCDLDOQSKPISLPEDLPG	439
Qy	440	ASYTLSOOCELAFGVGSKPCPYMQYCTKLWCCTGKAKGMVCOTRHPFNADGTSCEGKLC	499
Db	440	ASYTLSOOCELAFGVGSKPCPYMQYCTKLWCCTGKAKGMVCOTRHPFNADGTSCEGKLC	499
Qy	500	LKGACVERHNLNHRVDGWSWAKWDYPGPGSRTCCTGGGVQLARRQCTNPTTPANGGKYCEGVR	559
Db	500	LKGACVERHNLNHRVDGWSWAKWDYPGPGSRTCCTGGGVQLARRQCTNPTTPANGGKYCEGVR	559
Qy	560	VKYRSCNLEPCPSASGKSPREQCCEAFNGYHNHSTNRLTLAVAWVPKYSVGSVPRDKCKLI	619
Db	560	VKYRSCNLEPCPSASGKSPREQCCEAFNGYHNHSTNRLTLAVAWVPKYSVGSVPRDKCKLI	619
Qy	620	CRANGTGYFYVLAPKVVYDGTCLSPDSTSVCVQKCKIKAGCDGNLGSKKRFDKCGVCGGDN	679
Db	620	CRANGTGYFYVLAPKVVYDGTCLSPDSTSVCVQKCKIKAGCDGNLGSKKRFDKCGVCGGDN	679
Qy	680	KSCKKVYTGLTFTKPMHGYNFVVAIPAGASSIDIQRGYKGLIGDDNYLAKNSQKGYLLNG	739
Db	680	KSCKKVYTGLTFTKPMHGYNFVVAIPAGASSIDIQRGYKGLIGDDNYLAKNSQKGYLLNG	739
Qy	740	HFVVSAYERDLVKGSLRLYSGTGTAVESLQAASRPILPLETTVEVLSVGKMTPPRVRSFY	799
Db	740	HFVVSAYERDLVKGSLRLYSGTGTAVESLQAASRPILPLETTVEVLSVGKMTPPRVRSFY	799
Qy	800	LPKEPREDKSSHPRGDPVSLHNSVLSLSNQYVEQDDPPARWVAGSWGPCASCSGSL	859
Db	800	LPKEPREDKSSHPRGDPVSLHNSVLSLSNQYVEQDDPPARWVAGSWGPCASCSGSL	859
Qy	860	QKRAVDRCGSAGQRTVPACDAARHPVETQACGBCPTWELSAWSPCKSCGRGFRQRSLK	919
Db	860	QKRAVDRCGSAGQRTVPACDAARHPVETQACGBCPTWELSAWSPCKSCGRGFRQRSLK	919
Qy	920	CVHGGRLRLARDQCNLHRKKPOELDFCVLRPC	950
Db	920	CVHGGRLRLARDQCNLHRKKPOELDFCVLRPC	950

```

RESULT 2
US-10-163-316-2
; Sequence 2, Application US/10163316
; Publication No: US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor.
; FILE REFERENCE: MP101-023PIRNM
; CURRENT APPLICATION NUMBER: US/10/163.316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2

Query Match      42.9%; Score 408; DB 9; Length 823;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	107	VNAEPDSFAAVSLCGGLRGAFGYRGAEEYVISPLPNASAPAAQNSOGAHLQRRGYPGGP	166
Db	107	VNAEPDSFAAVSLCGGLRGAFGYRGAEEYVISPLPNASAPAAQNSOGAHLQRRGYPGGP	166
QY	167	SGDPTSRCGVASGWNPAILRALDPYKPRRAGFGESESRRSRSGRAKREVSIPRYVETLVVA	226
Db	167	SGDPTSRCGVASGWNPAILRALDPYKPRRAGFGESESRRSRSGRAKREVSIPRYVETLVVA	226
QY	227	DESMVKFHGADLEHYLLTLTAARLYRHPSILNPINIVVVKVLLLRDSDGPKVTGNAA	286
Db	227	DESMVKFHGADLEHYLLTLTAARLYRHPSILNPINIVVVKVLLLRDSDGPKVTGNAA	286
QY	287	LTLNFCAWQKLNKVKSDKHPEYWDTAILTRODLGCATCTDLGMADVGTMDCKRRSCS	346
Db	287	LTLNFCAWQKLNKVKSDKHPEYWDTAILTRODLGCATCTDLGMADVGTMDCKRRSCS	346
QY	347	VIEDDGLPSAFTTAHELGHVFNPHDNVKVCEEVFGKLRANHMMSPTLIQIDRANPWSAC	406
Db	347	VIEDDGLPSAFTTAHELGHVFNPHDNVKVCEEVFGKLRANHMMSPTLIQIDRANPWSAC	406
QY	407	SAAIITDFLDSGHGDCLLDQPSKPISLPEDLPGASYTLSSQOCELAFGVGSKPCPYMOYCT	466
Db	407	SAAIITDFLDSGHGDCLLDQPSKPISLPEDLPGASYTLSSQOCELAFGVGSKPCPYMOYCT	466
QY	467	KLWCTGKAKGMQVQCTRHFPWADGTCGEGKLCCLKGACVERHNLNKHR	514
Db	467	KLWCTGKAKGMQVQCTRHFPWADGTCGEGKLCCLKGACVERHNLNKHR	514

RESULT 3
 US-09-965-631-6
 ; Sequence 6, Application US/09965631
 ; Patent No. US20020115842A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fiddle, Carl Johan
 ; APPLICANT: Hilbun, Erin
 ; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
 ; FILE REFERENCE: LEX-0241-USA
 ; CURRENT APPLICATION NUMBER: US/09/965,631
 ; CURRENT FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: US 60/236,689
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 367
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-09-965-631-6

QY 320 DLGGATTCITLGMADVGTMCMDPKRSCSVIEDDGLPSAFTTAHELG 364
 Db 320 DLGGATTCITLGMADVGTMCMDPKRSCSVIEDDGLPSAFTTAHELG 364

RESULT 4

US-09-965-631-2
 ; Sequence 2, Application US/09965631
 ; Patent No. US20020115842A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fridde, Carl Johan
 ; APPLICANT: Hilbun, Erin
 ; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Encodin
 ; FILE REFERENCE: LEX-0241-USA
 ; CURRENT APPLICATION NUMBER: US/09/965,631
 ; CURRENT FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: US 60/236,689
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 321
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-09-965-631-2

Query Match 31.6%; Score 300; DB 10; Length 321;
 Best Local Similarity 100.0%; Pred. No. 4.4e-279;
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 EPERVVPVIRLDPDINGRRYWRGPDGQGLIFQITAFQEDFYHLTTPDAQFLAPAF 79
 Db 20 EPERVVPVIRLDPDINGRRYWRGPDGQGLIFQITAFQEDFYHLTTPDAQFLAPAF 79
 QY 80 STEHLGVPLQGTGSSDLRCFYSGDVNAEPDSPAAYSLCGGLRGAFYGAEYVISPL 139
 Db 80 STEHLGVPLQGTGSSDLRCFYSGDVNAEPDSPAAYSLCGGLRGAFYGAEYVISPL 139
 QY 140 PNASAPAAQNSQGAHLQVRGPGGSDPTSRGVSAGNPAILLRALDYPKRRAGFG 199
 Db 140 PNASAPAAQNSQGAHLQVRGPGGSDPTSRGVSAGNPAILLRALDYPKRRAGFG 199
 QY 200 ESRRRSRGRKRFYSIPRYVETLVVADESVMVKFHGADLEHLLTLATAARLYRHPISIL 259
 Db 200 ESRRRSRGRKRFYSIPRYVETLVVADESVMVKFHGADLEHLLTLATAARLYRHPISIL 259
 QY 260 NPINIVVVKVLLLRDSDGPKVTGNAALTNRNCAWQKKLNKVSQKHPEYWDTAILFTQ 319
 Db 260 NPINIVVVKVLLLRDSDGPKVTGNAALTNRNCAWQKKLNKVSQKHPEYWDTAILFTQ 319

RESULT 5

US-09-803-589-2
 ; Sequence 2, Application US/09803589
 ; Patent No. US20020112251A1
 ; GENERAL INFORMATION:
 ; APPLICANT: McCarthy, Sean A.
 ; APPLICANT: Holtzman, Douglas A.
 ; APPLICANT: Goodearl, Andrew D.J.
 ; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
 ; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
 ; FILE REFERENCE: 07334-325001
 ; CURRENT APPLICATION NUMBER: US/09/803,589
 ; CURRENT FILING DATE: 2001-03-09
 ; PRIOR APPLICATION NUMBER: US 09/128,709
 ; PRIOR FILING DATE: 1998-08-04
 ; PRIOR APPLICATION NUMBER: US 60/054,645
 ; PRIOR FILING DATE: 1997-08-04
 ; PRIOR APPLICATION NUMBER: US 09/130,491
 ; PRIOR FILING DATE: 1998-08-06
 ; PRIOR APPLICATION NUMBER: US 60/054,966

; PRIOR FILING DATE: 1997-08-06
 ; PRIOR APPLICATION NUMBER: US 60/058,108
 ; PRIOR FILING DATE: 1997-09-05
 ; PRIOR APPLICATION NUMBER: US 09/388,280
 ; PRIOR FILING DATE: 1999-09-01
 ; PRIOR APPLICATION NUMBER: US 09/388,279
 ; PRIOR FILING DATE: 1999-09-01
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 608
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-803-589-2

Query Match 1.8%; Score 17; DB 10; Length 608;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
 Db 91 AFTTAHELGHVFNMPHD 107

RESULT 6

US-09-803-589-8
 ; Sequence 8, Application US/09803589
 ; Patent No. US20020112251A1
 ; GENERAL INFORMATION:
 ; APPLICANT: McCarthy, Sean A.
 ; APPLICANT: Holtzman, Douglas A.
 ; APPLICANT: Goodearl, Andrew D.J.
 ; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
 ; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
 ; FILE REFERENCE: 07334-325001
 ; CURRENT APPLICATION NUMBER: US/09/803,589
 ; CURRENT FILING DATE: 2001-03-09
 ; PRIOR APPLICATION NUMBER: US 09/128,709
 ; PRIOR FILING DATE: 1998-08-04
 ; PRIOR APPLICATION NUMBER: US 60/054,645
 ; PRIOR FILING DATE: 1997-08-04
 ; PRIOR APPLICATION NUMBER: US 09/130,491
 ; PRIOR FILING DATE: 1998-08-06
 ; PRIOR APPLICATION NUMBER: US 60/054,966
 ; PRIOR FILING DATE: 1997-08-06
 ; PRIOR APPLICATION NUMBER: US 60/058,108
 ; PRIOR FILING DATE: 1997-09-05
 ; PRIOR APPLICATION NUMBER: US 09/388,280
 ; PRIOR FILING DATE: 1999-09-01
 ; PRIOR APPLICATION NUMBER: US 09/388,279
 ; PRIOR FILING DATE: 1999-09-01
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 608
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-09-803-589-8

Query Match 1.8%; Score 17; DB 10; Length 608;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
 Db 91 AFTTAHELGHVFNMPHD 107

RESULT 7

US-10-105-929-13
 ; Sequence 13, Application US/10105929
 ; Patent No. US20020137142A1

```
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/10/105,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-105-929-13

Query Match          1.8%; Score 17; DB 12; Length 608;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
DB 91 AFTTAHELGHVFNMPHD 107
|||||

RESULT 8
US-10-097-597-1
; Sequence 1, Application US/10097597
; Publication No. US20030022352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Ei-ji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: Pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-597-1

Query Match          1.8%; Score 17; DB 9; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
DB 156 AFTTAHELGHVFNMPHD 172
|||||

RESULT 9
US-10-097-597-12
; Sequence 12, Application US/10097597
; Publication No. US20030022352A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Ei-ji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: Pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-097-597-12

Query Match          1.8%; Score 17; DB 9; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
DB 156 AFTTAHELGHVFNMPHD 172
|||||

RESULT 10
US-10-097-580-1
; Sequence 1, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Ei-ji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-580-1

Query Match          1.8%; Score 17; DB 9; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
DB 156 AFTTAHELGHVFNMPHD 172
|||||
```



```
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-105-929-2

Query Match      1.8%; Score 17; DB 12; Length 967;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 356 AFTTAHELGHVFNMPHD 372
   |
Db 396 AFTTAHELGHVFNMPHD 412

Search completed: April 29, 2003, 17:25:35
Job time : 31 secs
```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 45.0027 seconds
(without alignments)
3378.970 Million cell updates/sec

Title: US-10-009-332-1_copy_213_950
Perfect score: 4043
Sequence: 1 FVSIPRYVETLVVADESWK.....DQCNLHRKQELDFCVLRPC 738

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.vertibrate.*
- 14: sp.unclassified.*
- 15: sp.virus.*
- 16: sp.bacteriap.*
- 17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4043	100.0	950	4 Q8TE58	Q8TE58 homo sapien
2	1685	41.7	340	11 Q91256	Q91256 mus musculus
3	1567	38.8	2165	5 Q19791	Q19791 caenorhabdi
4	1152	28.5	1688	5 Q8SX80	Q8SX80 drosophila
5	1143	28.3	1081	4 Q8TE60	Q8TE60 homo sapien
6	1139.5	28.2	1072	4 Q8TE57	Q8TE57 homo sapien
7	1119.5	27.7	1229	5 Q9VF61	Q9VF61 drosophila
8	1117	27.6	1054	5 Q9W493	Q9W493 drosophila
9	1057	26.1	1095	4 Q8TE56	Q8TE56 homo sapien
10	1024	25.3	1223	4 Q8WX58	Q8WX58 homo sapien
11	1022.5	25.3	1159	4 Q8TEY8	Q8TEY8 homo sapien
12	1020	25.2	1223	4 Q8TE55	Q8TE55 homo sapien
13	1012	25.0	1207	4 Q8TE59	Q8TE59 homo sapien
14	865	21.4	1427	4 Q96L37	Q96L37 homo sapien
15	830	20.5	269	6 Q9GL54	Q9GL54 oryctolagus
16	706	17.5	1280	11 Q9EPX2	Q9EPX2 mus musculus

17	693	17.1	1235	4	Q95428	Q95428 homo sapien
18	681	16.8	2167	5	Q76840	Q76840 caenorhabdi
19	663.5	16.4	1572	5	O44938	O44938 haemochus
20	581.5	14.4	2174	5	Q9G0R0	Q9G0R0 drosophila
21	579.5	14.3	3060	5	Q9VAV4	Q9VAV4 drosophila
22	569.5	14.1	766	4	P82987	P82987 homo sapien
23	558	13.8	192	6	Q95N24	Q95N24 equus caball
24	557	13.8	3198	5	Q9U8G8	Q9U8G8 manduca sex
25	553	13.7	1062	5	Q19204	Q19204 caenorhabdi
26	550.5	13.6	525	4	Q96RW4	Q96RW4 homo sapien
27	518	12.8	187	6	Q95N23	Q95N23 equus caball
28	510.5	12.6	872	5	Q22580	Q22580 caenorhabdi
29	504	12.5	1444	5	Q17591	Q17591 caenorhabdi
30	493	12.2	951	4	O60345	O60345 homo sapien
31	487.5	12.1	1014	5	Q95R33	Q95R33 drosophila
32	429	10.6	1091	5	Q9W126	Q9W126 drosophila
33	426	10.5	790	5	O8T458	O8T458 drosophila
34	411.5	10.2	1059	5	P90884	P90884 caenorhabdi
35	410	10.1	183	11	Q99JP1	Q99JP1 mus musculus
36	402.5	10.0	364	4	Q9UGQ1	Q9UGQ1 homo sapien
37	343	8.5	622	4	Q9H8X0	Q9H8X0 homo sapien
38	338	8.4	192	11	Q9CX59	Q9CX59 mus musculus
39	332.5	8.2	105	6	Q9GLK7	Q9GLK7 oryctolagus
40	321.5	8.0	244	11	Q9CRC7	Q9CRC7 mus musculus
41	303	7.5	108	6	Q9GLK6	Q9GLK6 oryctolagus
42	302.5	7.5	1637	6	Q9XSU8	Q9XSU8 bos taurus
43	302.5	7.5	5146	6	Q8SPM4	Q8SPM4 bos taurus
44	298	7.4	117	11	O8VHK4	O8VHK4 mus musculus
45	279.5	6.9	1081	5	Q9U631	Q9U631 drosophila

ALIGNMENTS

RESULT 1

Q8TE58 PRELIMINARY; PRT; 950 AA.
 ID Q8TE58
 AC Q8TE58;
 DT 01-JUN-2002 (TRENBLrel. 21, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Metalloprotease disintegrin 15 with thrombospondin domains.
 GN ADAMTS15.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21856482; PubMed=11867212;
 RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
 RA Lopez-Otin C.;
 RT "Cloning, expression analysis, and structural characterization of
 RT seven novel human ADAMTSs, a family of metalloproteinases with
 RT disintegrin and thrombospondin-1 domains.*";
 RL Gene 283:49-62(2002).
 DR EMBL; AJ315733; CAC86014.1; -;
 KW Integrin; Protease.
 SQ SEQUENCE 950 AA; 103286 MW; 5DFBE18285CCCC3B CRC64;

Query Match 100.0%; Score 4043; DB 4; Length 950;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVSIPRYVETLVVADESWKFGADLEHYLTLLATAARLYRHPSILNPINVVVKVLL 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 213 FVSIPRYVETLVVADESWKFGADLEHYLTLLATAARLYRHPSILNPINVVVKVLL 272
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 61 RDRDSGPKVTGNAALTNRNFCQWOKLNKVSQKHPYWDTAILFTRODLGATTCDFLGM 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 273 RDRDSGPKVTGNAALTNRNFCQWOKLNKVSQKHPYWDTAILFTRODLGATTCDFLGM 332
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 121 ADVGTMCDFKSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKCEEVFGKLRANHWMP 180

```
Db 333 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVYKCEEVFGKLRANHWSP 392
QY 181 TLIQIDRANPWSACAAIITDPLDSDHGDCLLDQPSKPISELPEDLPASYYTLQOCELA 240
Db 393 TLIQIDRANPWSACAAIITDPLDSDHGDCLLDQPSKPISELPEDLPASYYTLQOCELA 452
QY 241 GVGSRPCPTMOCYCTKLWCTGKAGQWVCQTRHFWADGTSCEGKGLCLKGACVERHNLN 300
Db 453 GVGSRPCPTMOCYCTKLWCTGKAGQWVCQTRHFWADGTSCEGKGLCLKGACVERHNLN 512
QY 301 HRVDGSAWAKWDYPCSRTRCCTGGGVOLARRQCTNPTPANGKYCGVRVYKRSCLNLEPCPS 360
Db 513 HRVDGSAWAKWDYPCSRTRCCTGGGVOLARRQCTNPTPANGKYCGVRVYKRSCLNLEPCPS 572
QY 361 SASGKSFFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKLKLICRANGTGYFYVLA 420
Db 573 SASGKSFFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKLKLICRANGTGYFYVLA 632
QY 421 PKVVDGTLCSPDSTSVCVQKCIKAGCDNGLGSKRRPKCGVCGDNKSKCKVTGLFTKP 480
Db 633 PKVVDGTLCSPDSTSVCVQKCIKAGCDNGLGSKRRPKCGVCGDNKSKCKVTGLFTKP 692
QY 481 MHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNGHFVYVSAVERDLV 540
Db 693 MHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNGHFVYVSAVERDLV 752
QY 541 KGSLLRYSGTGTAVESLOASRPILPLAVEVLSVGMKTPPRVYSFYLKPEPREDKSSHP 600
Db 753 KGSLLRYSGTGTAVESLOASRPILPLAVEVLSVGMKTPPRVYSFYLKPEPREDKSSHP 812
QY 601 KDPGSPVLHNSVLSNQVEQDPPARVAVAGSWGPCSCASGSGLOKRAVDCRGSAQ 660
Db 813 KDPGSPVLHNSVLSNQVEQDPPARVAVAGSWGPCSCASGSGLOKRAVDCRGSAQ 872
QY 661 RTVPACDAARHVPETQACGECPTWELSAWSPCKSCSGRGFQRLSKVCGHGGRLARQ 720
Db 873 RTVPACDAARHVPETQACGECPTWELSAWSPCKSCSGRGFQRLSKVCGHGGRLARQ 932
QY 721 CNLHRKPQELDFCVLRPC 738
Db 933 CNLHRKPQELDFCVLRPC 950

RESULT 2
Q91256 PRELIMINARY; PRT: 340 AA.
AC Q91256;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE Similar to a disintegrin and metalloproteinase with thrombospondin
DE motifs 1 (ADAMTS-1) (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009667; AA09667.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR NON_TER
FT INTEGRIN.
SQ SEQUENCE 340 AA; 37021 MW; 06562C747634A8BD CRC64;

Query Match 41.7%; Score 1685; DB 11; Length 340;
Best Local Similarity 91.7%; Pred. No. 4.2e-147;
Matches 308; Conservative 11; Mismatches 17; Indels 0; Gaps 0;

QY 403 KCKLICRANGTGYFYVLAAPKVVVDGTLCPDSTSVCVQKCIKAGCDNGLGSKRFRKCGV 462
```

```
Db 5 KCKLICRANGTGYFYVLAAPKVVVDGTLCPDSTSVCVQKCIKAGCDNGLGSKRFRKCGV 64
QY 463 CGGDNKSCKKVTGLTKPMHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGR 522
Db 65 CGGDNKSCKKVTGLTKPMHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGR 124
QY 523 YLLNGHFVYVSAVERDLVYKGLRYSCTGTAVESLOASRPILPLAVEVLSVGMKTPPRV 582
Db 125 YLLNGHFVYVSAVERDLVYKGLRYSCTGTAVESLOASRPILPLAVEVLSVGMKTPPRV 184
QY 583 RYSFYLKPEPREDKSSHPKDPGRGSLVLSNSQVEQDPPARVAVAGSWGPCSCAS 642
Db 185 RYSFYLKPEPREDKSSHPKDPGRGSLVLSNSQVEQDPPARVAVAGSWGPCSCAS 244
QY 643 CGSGLOKRAVDCRGSAQRTVPACDAARHVPETQACGECPTWELSAWSPCKSCSGRGFQ 702
Db 245 CGSGLOKRAVDCRGSAQRTVPACDAARHVPETQACGECPTWELSAWSPCKSCSGRGFQ 304
QY 703 RRSCLKVCGHGGRLARQDQCLRRKFPQELDFCVLRPC 738
Db 305 RRSCLKVCGHGGRLARQDQCLRRKFPQELDFCVLRPC 340

RESULT 3
Q19791 PRELIMINARY; PRT: 2165 AA.
ID Q19791;
AC Q19791; Q27524;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gajadaty S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1; -.
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1; -.
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSSP; P15167; 1DTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 14.
DR SMART; SM00209; TSP1; 18.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;
```


Db 637 SQKCPDSDS--VDFAAQAQAEHNSRRFRGRHYK-----WKP-YTQVEDQDLCKLCYCI 684
QY 410 ANGTGYFYVLAPKVVDGTLCSDPDSTSVQVQKCIKAGCDNGLSKKREKDCGCGGDNKS 469
Db 685 AEGFDFFFSLSNKVKDGTCSDESRNVCIDGICERVCNDVLGSDADEVCGVCGNNSA 744
QY 470 CKKVTGLTKPMH--GYNFWVAIPAGASSIDIRQRYGKLGIDDDNYLALKNSQGYKLLNG 527
Db 745 CTIHRGLYKHHHTNOYHYHMVTPSGARSIRIYEMNV-----STSVISVRNALRYLNG 799
QY 528 HFVSAVERDLVVKSGLLRYSTGTAVESLQASRPILLETVEVLSVGKMTPPRVRYSPY 587
Db 800 HWTVDWPGR-YKFSGTYTFRYSYNEPENLIATGPTNETLIVELLFQGR--NPGVAWEYS 856
QY 588 LPKEPREDRKSHPKDPRGPS-----VLHNSVLSLSNOVE---QP 623
Db 857 MPR-----LGTEKQPPAOPSYTWAIVRSKSVSCGGQMTVREGCYRDLKFOVNMFCNP 911
QY 624 DRDP-----PARWVAGSWGPCASCGSLQKRAVDC--RGSAGQRTVPA--CD 667
Db 912 KTRPVTGLVPCKVSACPPSPWSVGNNSACSRCTCGGAQSRPVQCTRRVHYDSEPVFASLCP 971
QY 668 AARHPVETQAC--GEP--PTWELSAWSPCKSCGRGFQBSRLKCVHG-----GRLLARQOC 721
Db 972 QP-APSSRQACNSQCPPAWSAGPNAECSTCGKWRKRAVACKSTNPSSARAQLLPDAVC 1030
QY 722 NLHRKPQELDFCVLRPC 738
Db 1031 TSEPMPRMEACLLQRC 1047
RESULT 7
QYVF61 ID QYVF61 PRELIMINARY; PRT; 1229 AA.
AC QYVF61;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CG6107 protein.
GN CG6107.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eurygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,

Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL; AE003709; AAF55199.1; --
DR FlyBase; FBgn0038340; CG6107.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 2.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS00215; ADAM_MEPRO; 2.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; zinc.
SQ SEQUENCE 1229 AA; 138136 MW; C1293180AB19B700 CRC64;
Query Match 27.7%; Score 1119.5; DB 5; Length 1229;
Best Local Similarity 34.5%; Pred. No. 4.9e-94;
Matches 281; Conservative 111; Mismatches 260; Indels 163; Gaps 35;
QY 8 VETLVWADESXVPHGADLEHLLTLATAARLYRHPISILNPINIVVVKVLLLRDRDSGP 67
Db 356 LEVLIAVDNSMKQFHGDLQPIYILMSIVSSIFADASIGNSIRILLVRLISL-----P 409
QY 68 KV---TGNAAALTANFCAWQKLNKVSXDKHPEYWDTAIFTRQDLGAT---TCDPLGMA 121
Db 410 NINDQTHSSNEMLKHFQCF---INQGYER---DTAMLITREPICGSPVGKICHMLGLA 462
QY 122 DVGTMCDPKRSVIEDDGLPSAFTTAHELCHVFNPMDNVKVCYEEVFGKLRAN---HMM 178
Db 463 ELGTVCSSSSCSIVQDTGLPTAFAHELGHILNNHDDDDCKMYPYVTRQNNKVLHIM 521
QY 179 SPTL-IOIDRANPWSACSAIITDLDLSDHGDCLLPQSPKPSLP---EDLPQASYTLQ 234
Db 522 SSVNGIHM---HPWSWSKK---TD-----KSCL--ETSVGAHIPYGTERTLPGEIYSLDA 567
QY 235 QCELATGCVGSKPCPYMOCYCTKLNC---TGKAKQMVQCTRHFPWADGTSCG--EGKLCILKG 290
Db 568 QCQLSFQNDRGYCTPTDEECARLNCNRTSGNSNEQ---CASNLPWADGTPCGSSGHWCRG 625
QY 291 ACYVRHNLNKH---RVDSGSAWKDPYPCSRTRTCGGGVQLARRQCTNPPTPANGKCYCEGV 346
Db 626 KCVS---NKHGYGRQVNGWGWPPTPTPCSLTCGGVQESRECNQPPENGKKYCTGS 681
QY 347 RVKYRSCNLEPCPSSASGSKSFREOCEAFNGYHNHNRILTAVAWPKYSGVSPRDKCKL 406
Db 682 RKIYRSCNTHQCPGSGNDP--REQQYAMNRMNIPGVNPDTKWPKYE---KDACKL 735
QY 407 ICRANGTYFVYLAPKVDGTLCSDPDSTSVQVQKCIKAGCDNGLSKKREKDCGCGCD 466
Db 736 FCRMDKMYTFMLKSMVTDGTSKAVDSFKCVNGICRPAGCNELNSIAKLDKCGVCEGR 795
QY 467 NKSCKKVTG-LFTKPMHGYN-----FVVAIPAGASSIDIRQRYGKLGIDDDNYLA 515
Db 796 NDTCHVEVTGNLLNLLGLNDGNPNKTLVYVTRIPKASNIITQGYP---DQNFIV 851
QY 516 LKNSQGYLLNGHFVYSAVERDLVWKGSLRLYSGTGTAVESLQA---SRILPLPLVVELS 573
Db 852 LTDDRDELNGKFL-LKTYPLKPYAGVTMQYTGSSSVVEQVNTTYSWLSRDLIVQIIS 910

Qy 574 VGKMTPPR-----VRYSYFLPKPPRE----- 594
Db 911 LDVSPSKRQDTLLSYSTYIDKPDYAEVEIYRWEMQAPSNCDLSLGRSHRLPACIST 970
Qy 595 -----DKSSHPK-DPRGPSV---LHNSVLSLSNOVEQPDPPARWAGSWGPC 639
Db 971 TQGVKVAPOFCDSKAMPKIDDRACNTDCLRLNLTVTSISE-----C 1010
Qy 640 SASGCS-GLQKRAVDC--RGSAGRT---VPACDRAHRPVEVTAQGECPPTWELSAWSP 692
Db 1011 SAACGELGTREKTVACVQTFTNMQRSNIVDMYSCKLKFDVAYHEECREGC--WVLSWST 1068
Qy 693 CSKSCGRGFRRLSKCVGHGRLRLAQCNLHRKP 727
Db 1069 CSKSCGTGSGQREAHCVLHNSR-VSDCLCNPRTKP 1102
RESULT 8
Q9W493
ID Q9W493 PRELIMINARY; PRT: 1054 AA.
AC Q9W493;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CG4096 protein.
GN CG4096.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Culey S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.-H., Zhong F.-N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;

RL Science 287:2185-2195(2000).
DR EMBL; AE003435; AAF46065.1; --
DR HSP; P15167; IATL.
DR MEROPS; M12.231; --
DR FlyBase; FBgn0029791; CG4096.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_Mrptptdse.
DR Pfam; PF015621; Pep_M12B_propep; 1.
DR Pfam; PF00090; tsp-1; 2.
DR SMART; SM00209; TSPI; 2.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPI; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1054 AA; 118616 MW; DC15455555CB6212 CRG64;
Query Match 27.6%; Score 1117; DB 5; Length 1054;
Best Local Similarity 35.0%; Pred. No. 6.6e-94;
Matches 276; Conservative 116; Mismatches 276; Indels 120; Gaps 34;
Qy 2 VSIPRYVETLVVAVDESMVVFHGADLEHYLLTLATAARLYRHPISLNPINIVVVKVLLLR 61
Db 321 ISSPRHVEVLIVADATMSAFH-RDLNGYLLTIMNVVSALYKPSIGNSIEIVVRIQLD 379
Qy 62 DRDSGPK--VTGNAALTLRNFCAWOKKLNKVSUKHPEYWDTAIFLTRQDLGATTCDTLG 119
Db 380 EESQLQLNLTQNAQKNLDRFCWQHKLKNGSEKDPHHHDVAILLTRKNIC-ANNCMTLG 438
Qy 120 MADVGTWCDPKSCSVIEDDGLPSAFTTAHGLGHVFNPHDNKV-CEEVFCKLRANIMM 178
Db 439 LANVGMCKPKSCSVYEDNGIMLSHTITHELGHVFMHFDATKIGCHPRVGPV--VHIM 496
Qy 179 SPTL----TIDRANPWSACSAITDFLDSGHGDCLLDQPSKPI---SLPEDLPASVT 231
Db 497 TPTFGADITLV---CWSNCSRKTYTHFLDQGLGEC-LDDPTPLDEYNTGELPGMYN 551
Qy 232 LSQOCELAFG-----VSGKPCPYMOYCTKLMCTKAKQMCQVQTHFFWADGTSCEGEG 285
Db 552 ARGQCRQLFNLTATDSEVGACSAFH-EFCSTLWC--KVNGE--CVTHMRPTAPGTLGCRNK 606
Qy 286 LCLKGACVSRHNLKRVHGVSNKAWDPYGCSTCGGVOLARROCTNTPANGKYKCG 345
Db 607 WCQNGKCVREEL--AAVNGMGWDSEWSECSKSGGGVSTQORECDNVPANGVFCIG 664
Qy 346 VRKYRSCNLEPCPSSASGSKSFEEQCEAFN--GYNHSTNRLTLAVAWVPKYSGVSPDK 403
Db 665 ERK3YKICRKRCP--AEPSPFAQOCARFNVSYOGATYK-----WLPFFDKNNP--- 713
Qy 404 CKLICRANGTGYVYVLAQKVVDTGLSPDSTVCVCGKCIKAGCDGNLGSKRFRKCGYC 463
Db 714 CKLFCSDVDDTITANNGATVLDGTCTLTNNNCIDGICKKVGCDWIDVSEVODDCGVC 773
Qy 464 GGDNCKCKVTGLFTKPM---HGYNFVVAIPAGASSIDIRORYKGLIGDNNYLAKNS 519
Db 774 GGSQDQCPVREYITDFFRAKDGAYVEITIPARAHILIRE-----LANSHPFLAATG 828
Qy 520 QG---KYLINGHFVYSAVERDLVVVGLSLRYSGTGTAVESLQASRPILPLEPLTEVLSVGM 577
Db 829 DGGDRFYLNGDSLIS-MPGEFEIAGAESLYDRVDEQ-ETITIPQTIQHSISLYAIVRGNE 886
Qy 578 TTPRVRYSYFLPREDKSSHPKDPGRPSVLSLSNQVEQPDPPARWAGSWG 637
Db 887 SNAG3IEYFETLP-----ALN---VTAGROFQ-----WRLSNWT 916
Qy 638 PCSASGSGGLQKRAVDC---RGSAGQ-----RTVPACDAA---HRPV- 673
Db 917 ACSASGGGVQVHREPCQENGKESNEPQRIYSIKRNLIIALGDTLPCWTHAKNKRPAR 976
Qy 674 ETQACQ-EPQPT-WELSAWSPCSKSGCR-GF-----QRRSLKCVGHGRLRLAQCNLHR 725
Db 977 QSRGCGGCPCPAHWWPGPMQFCVPTCPVGFVAPPORRRSVVCLDEHVDVVADEAGHLQ 1036

```
QY 726 KPOELDFC 733
Db 1037 KPAEMEPC 1044

RESULT 9
Q8TE56 PRELIMINARY; PRT; 1095 AA.
AC Q8TE56;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Metalloprotease disintegrin 17, with thrombospondin domains.
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315735; CAC86016.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1095 AA; 121099 MW; A5C6540484754D5D CRC64;

Query Match 26.1%; Score 1057; DB 4; Length 1095;
Best Local Similarity 32.6%; Pred. No. 2.5e-86;
Matches 261; Conservative 109; Mismatches 290; Indels 140; Gaps 27;

QY 8 VETLVADSVKVFHAGD-LEHYLLTLTATAARLYRHPHSILNPINIVVKKVLLLRDRSG 66
Db 234 VETLVADSVKVFHAGD-LEHYLLTLTATAARLYRHPHSILNPINIVVKKVLLLRDRSG 66
QY 67 PKVTGNAALTLRNFCAWQK-----LNKV--SDKHPEYDTAILFTRODLC--GAT 113
Db 294 LSIHHGERSLESCHWQNEEYGGARYLGNQVGGKDDPLVDAAVFTYKDFCVKDE 353
QY 114 TCDILGMADVTCMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKKVCEVFGKLR 173
Db 354 PCDTIVGIAYLGGVCSAKRKCVLAEENGLNLAFTIAHELGHNLGMNH-----DDHSSCAG 408
QY 174 ANHMSPTLIQIDRANP-----WSACSAAILTDFLDSHGDCLL---DQPSKPTSLPEDLP 226
Db 409 RSHMSGWVK--GRNPDDLWSKSRDLENFLKSVSTCLLVTDPRSQHTVLPKPLP 466
QY 227 GASTYLSQOCBELAFGVGSKPCPYMQY--CTKLWCTGKAKGMQVCTRHFWADTSCGEG 284
Db 467 GMHYSANEQCOILFGMNATFCRNHEHLMCAGLWCL--VEGDTSCKTKLDPLDTEGCAD 524
QY 285 KLCLKGACVERHNLKRVDSWAKWDPYGPCSTCGGVQLARRQCTNTPANGKYCE 344
Db 525 KWCRAECVSTPIPEH--VDGDWSPWAGWSMCSRTCTGTFARFRQRKCDNPPPGPGTHCP 583
QY 345 GVRKYRSCNLEPCPSSAGSKSFREOCEAFNGYNHSTNRLTLAVANPKYSGVSPDKC 404
Db 584 GASVEHAVCNLPCPKGL--PSFRDQCOAHDRLLSPKKGLLTA-----VVDKPC 633
QY 405 KLICRANGTGYVYLAAPKVDGTLCSPDSTSVCVQGVKICAKCDGNLGRKRFDKCGVC 464
Db 634 ELYCSPLGKESPLLVADRVLDGTPCGPYETDLCVHGKCGKIGCDGIIIGSAKEDRCGVCS 693
QY 465 GDNKSKVTVGLFTKPMHGVNFVVAIPAGASSIDIRGKYGKLGIDGNYLALNKSQKYL 524
Db 694 GDGKTCMLVKGDFSH-----ARGTALKDSG--KGSI-----NSDWKIE 729
QY 525 LNHGFVVAVERDLVWKGSLRLRYSGTGTAVESLQASRPILPTEVL-----SVGKWT 579
```

```
Db 730 LPGEFOIA-----GTTVRYVRRG-LWEKISAKGPTKLPLHLMLVLLFHDQDYG---- 775
QY 580 PRVRYSEYLPKPREDKSSHPKDPKPR-----GPSV-----LHNS 612
Db 776 --IHVEYTVPNRTAENQSEPEKPDQSLFIWTHSGWEGSCVQCGGERRIVSCTRIVNK 833
QY 613 VLSLSNQVEQDDRRPP-----ARWAGSWGPCSCASGSGLOKRAYDC-----R 655
Db 834 TTTLVNDSDCPQASRPEFQVRRRCNLHPCQSRWVAGWSPSCATCEKGFQHVRETVCVQLQ 893
QY 656 GSAGORTVPACDAAHRPVETQAC-GEPC-PTWELSAWSPCKSCGRCGFQRRSLKCVGHGG 713
Db 894 NGTHVTRPLCYCPGPRPAAVQSCGQDCLSIWEASEWSQCSASCGKGVKRTVACTNSQG 953
QY 714 RLARDQCNLHRKPOELDFC 733
Db 954 K-----CDASTRPRAEAC 967

RESULT 10
Q8WX58 PRELIMINARY; PRT; 1223 AA.
AC Q8WX58;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE A disintegrin-like and metalloprotease with thrombospondin type 1
DE motif 14 precursor.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638061; PubMed=11779638;
RA Bolz H., Ramirez A., von Brederlow B., Kubisch C.;
RT "Characterization of ADAMTS14, a novel member of the ADAMTS
RT metalloproteinase family.";
RL Biochim. Biophys. Acta 1522:221-225(2001).
DR EMBL; AF358666; AAL40229.1; -.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00090; TSP-1; 4.
DR PROSITE; PS00209; TSP1; 4.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR SIGNAL; Integrin; Protease; Metalloprotease.
FT SIGNAL 1 22
SQ SEQUENCE 1223 AA; 133871 MW; 36394AC4D92F170F CRC64;

Query Match 25.3%; Score 1024; DB 4; Length 1223;
Best Local Similarity 31.7%; Pred. No. 3.2e-85;
Matches 265; Conservative 95; Mismatches 296; Indels 180; Gaps 32;

QY 8 VETLVADSVKVFHAGD-LEHYLLTLTATAARLYRHPHSILNPINIVVKKVLLLRDRSG 66
Db 261 IEVLLVDDSVKVFHAGD-LEHYLLTLTATAARLYRHPHSILNPINIVVKKVLLLRDRSG 66
QY 67 PKV-TGNAALTLRNFCAWQKLNKVKSDKHPEYDTAILFTRODLCGATCTDLGMADVCT 125
Db 321 SLIERGNPSRLEQVCRWAHSQORQDPDPAEHHDHVVFLTRQDF-----GPSGYAPVTG 374
QY 126 MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHD---NVKVCVEVFGKLRANHMSPTL 182
Db 375 MCHPLRSCALNHEDGSSAFVIAHETGVLGMEHDGOGNCGADETSLGS-----VMAPLV 429
QY 183 IQIDRANPWSACSAAILTDFLDSHGDCDLDQPSKPI-SLPEDLPGASYTSLSQCELAG 241
```



```
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ345098; CAC87943.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1223 AA; 133925 MW; D585B6593977ED15 CRC64;

Query Match      25.2%; Score 1020; DB 4; Length 1223;
Best Local Similarity 31.6%; Pred. No. 7.6e-85;
Matches 264; Conservative 96; Mismatches 296; Indels 180; Gaps 32;

QY 8 VETLVVADSVKFKHGAD-LEHYLLTLTATAARLYRHPHSILNPINIVVVKVLLLRDRDSG 66
DB 261 IEVLLVVDVSVVRPHGKHQVNYLTMNIVDEIYHDESLGHNIALYRLINLVGYRQSL 320
QY 67 PKV-TGNALTLRNFCAKQKLNKVDKHPYWDTAILFTRODLCGATCTDTLGMADVGT 125
DB 321 SLIERGNPSRLEQVCRWAHSQQODPSHAHHVFLTRQDF-----GPSGYAPVTG 374
QY 126 MCDPKRSVIEDDGLPSAFTTAHELGHVFNPHD---NVKVCVEVFGKLRANHMSPTL 182
DB 375 MCHPLRSALNHEDGFSAFIAHETGHVLCMEHDGCGCAGDTSLS-----VMAPLV 429
QY 183 IQIDRANPWSACSAIAITDFDLSHGDCLLDQPSKPI-SLPEDLPASYYTISQOCELAFG 241
DB 430 QAAPFRHFWRSCSLELSRYLPS--YDCLLDPDFDPAWPPPELPGINYSWDEQCRDFG 487
QY 242 VGSKPC---PYMVCYTKLWCKGAKGQMVQCOTRHFPHADGTSCEGKLCILKGCVERHNL 298
DB 488 SGYCTLAFTFECKQLWCS-HPDNPFYCKTKKGPPLDGTETCAPGKCFKGHCIMKSPE 546
QY 299 NKHRVDSWAKWDYPGCSRTCCGGVQLARRQCTNPTPANGKYCEGVVRYKRSCLNCP 358
DB 547 QTYGODGWSWTKFGSCRSRSGGVSRSRSCNPNPAGYGRCLGPMFYQVCNSEEC 606
QY 359 PSSASGSKFREQCEAFNGY-----NHSTNRLTLAVAVVPGYSGVSPRDKCLICRANG 412
DB 607 PGTY--EDFAQQCAKRNYSYVHONAKHS-----WVP-YEPDDDAQKCELIQCSAD 654
QY 413 TGYFYVLAPKVVDDTLCG-PDSTSVCGKCIKAGCDNGLSKRDKPCGVCGDNKSCK 471
DB 655 TGDVVFNNQVVDGTRCSYRDPYSVCARGCEVPVGCCKEYSGMKADDKCGVCGDNSHC 714
QY 472 KVTGLTK--PMHGYNFVVAIPAGASSIDIRQRYKGLIGDDNYLALKNS-QGKYLIN-- 526
DB 715 TVKGTGKASKQAGALKLVQIPAGARHIQI-----EALKSPHRIYVKNQVTSFILNPK 769
QY 527 GHEVVSVERDLVVKGSILRAYSGTGTAVESLQASRPILPLETVESLVGKMTPPRVRYSF 586
DB 770 GK---EATSRFTTAMG-LEWEDAVEDAKESLKTSGPLPEAIALAL----- 811
QY 587 YLPKEPREDSKSHPKDPGRPS-----VLHNSVLSD--SNQV--EOPDDRPAPRWAGS 635
DB 812 -----PPTGEGPRSSLAYKYVTHEDLLPLIGSNVLLLEMD---TYEWALKS 855
QY 636 WGPCSASCGSLQKRAVDCR----- 655
DB 856 WAPCSKACGGGIRFTKYGCRRRDRHHVHRHLCDHKKRPKPIRRRSNQHPQPCSPVWVTEE 915
QY 656 -----GSAGORT-----VPACDAARHPVETQAG-----EPCPT-WELS 688
DB 916 WGACRSRYGKLVQTRIGIOCLPLSLNGTHKVPAPKACAGDRPBARPCLRVPCPAQWRLG 975
QY 689 AWSPCSKSCGGRGQRSLKC-----VHGGRLLARDQCNLHRKQELDFCVLRPC 738
DB 976 AWSQCATCGEGIQQRQVVRTNANSLGH-----CEGDR-PDTVQVCSLPAC 1021

RESULT 13
Q8TE59
ID Q8TE59 PRELIMINARY; PRT; 1207 AA.
```

```
AC Q8TE59;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE ADAMTS-19.
GN ADAMTS19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ311904; CAC84565.1; -.
SQ SEQUENCE 1207 AA; 134061 MW; AF36F6BF5886FDE2 CRC64;

Query Match      25.0%; Score 1012; DB 4; Length 1207;
Best Local Similarity 31.9%; Pred. No. 4.1e-84;
Matches 259; Conservative 117; Mismatches 295; Indels 140; Gaps 33;

QY 8 VETLVVADSVKFKHGAD-LEHYLLTLTATAARLYRHPHSILNPINIVVVKVLLLRDRDSG 66
DB 327 IETVVVADPAMVSYHGADAARRFILTMNVFLFOHKSGLGVQVNLVRVLIKLHETPPE 386
QY 67 PKVTGNAALTLRNFCAWQ-KLNKVSVDKHPEY---W-----DTAFLTRODLC--GAT 113
DB 387 LYIGHGHEKMLSEFCWKQHEEFCKNDIHELMSTNWGEDMTSDVAAILTRKDFCVHKDE 446
QY 114 TCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDN-VKYCEEVFGKL 172
DB 447 PCDTVGIAYLSGMCSEKRCIIAEDNGLNLAFTIAHEMHNHNGINHDNDHPSCAD---- 501
QY 173 RANHMSPPTLIQIDRAN--PWSACSAIITDFDLSHGDCLLD---QPSKPISLPEDLP 227
DB 502 -GLHIMSGEWIYGQNLGDSVSKSCKEDLERFLRSKASNCILQTNPQSVNSVMVPSKLP 560
QY 228 ASYTLSSQCELAFLGVSCKPCPYMNY--CTKLWCTGKAGQMVQCOTRHFPHADGTSCEGK 285
DB 561 MVTYADEQCOQLPLASFCQEQHVICHGLWC--KVEGEKCRKLPDPMQDGTDCDGLK 618
QY 286 LCLKAGCAVERHNLNHRVDGSHAKWDYPGCSRTCCGGVQLARRQCTNPTPANGKYCEG 345
DB 619 WCKAGECTSRISAPEH-LAGEWSLW---SPCSRTCSAGISSRERKC--PGLDSEARDNG 672
QY 346 VRVKYRSCNLEPCPSASGSKSFREQCEAFNGYNHSTNRLTLAVAVVPGYSGVSPDK-C 404
DB 673 PRKQYRICENPCPAGLPG--FRDQCCQAYSVRTSPKHIL-----QWQAVLDEEKPC 723
QY 405 KLICRANGTGYFYVLAPKVVDGTLCSPDSTSYCVGKCIKAGCDNGLSKRDKFCGVC 464
DB 724 ALFCSPVPGKEQPIILLSEKVMDCGTCGCGYGLDICANGRCQKVCYCDGLLSLAREDC 783
QY 465 GDNKSKCKVTGLFTFPM-HGYNFVVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQKY 523
DB 784 GNGKSKCIKIDFNHTRGAGYVEVLVIPAGARRIKVVEEK-----PAHSYLALRDA-GK 837
QY 524 LNLGHVVSVERDLVVKGSLLRYSG---TGTA--ESLQASRPILPLETVETV 571
DB 838 SINSWKI-----EHSGAPNLAGTVHYVYRGLWEKISAKGPTTAPLHLLV 883
QY 572 L-----SVGKMTPPRVRYSFYLPKEP-REDKSSHPKDP-----RG 605
DB 884 LLFQDQNYG-----LHVEYTIQSPDLPENOSKRAPELFMWTHTSWEDCDATCGGERK 937
QY 606 PSVLHNSVLISL-----NOVEQPDPRP-PARWAGSGWPCSGSGSLG 648
DB 938 TTVSCTKIMSKNISIVDNEKCKYLTKPEQIRKCNQPCQTRMMTWETPTCSRCTCGGMQ 997
```

372	RSVLVELPIAAVHGRWSWGPRSCRSRSCGGGVVTRRQCNPNRPAFGACRACVADLQAE	433
352	SCNLEPCPSSASGSKSFREEQAEFG-----YHNSTNRLTLAVAM---VPKYSG	397
432	MCNTQACEKIQ--LEFMSQCARTDGOPLRSSPGGASFVH-----WGAAVPHSQG	479
398	VSPEDKCLKLTRANGTYFYVLAPKVVDGTLCSF-----DST-SVCVGKCIKAGCDGNL	451
480	DA---LCRHMCRATGESFIMKRGDSFLDGTGRCPSPGREDGTLUSLCSGSCRTFGCDGRM	536
452	GSKRFRCKGCVGGGDNKSKCKVTGLFT--KPMHGYNFVVAIPAGASSIDIRQRYKGLIG	509
537	DSQQWDRQCVCVGGDNCSTSPRKGSFTAGRAREVFLVTVP-NLTSVVI--ANHRPLF-	592
510	DDNYLALKNSQGYLLNGHFVWS--AVERDLVKGSL-LRYSGTGTAIVSLQASR---PI	563
593	--THLAVRIG-GRYVWAGKMSISPTNTTTPSLLEDGRVEYRVALTEDRLPRLESIIRWGPL	649
564	LEPLTVEL-----SVCKMTPPRVRYSFYLPKPREDKSSHDPKDPGSPVLHNSVLSLSN	618
650	QEDADIQYRRYKGEYGNLTPDITFYFQPK-----	681
619	QVEQPDPRPARWY-AGSWGPCSCSGSLQKRAVDCRGSAGORTVPA--CDAHR-PVE	674
682	-----PRQAWYAAVARGPCSVSCGAGLRWNYSCLDQARKELVETVQCQSGSQPPAW	733
675	TOACG-EPCPT-WELSAWSPCSKSCGRGFORRSIKCVGHGGRLL-----ARDONLHRKP	727
734	PEACVLEPCPYWAVGDFPCSCSGGLRLRPRVRCVEAOGSLKTLPPARCRAQAQPA	793
728	QELDFCVLRPC 738	
794	VALETCNPQC 804	
RESULT 15		
Q9GL54	PRELIMINARY; PRT; 269 AA.	
AC	Q9GL54;	
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	Aggrecanase-2 (Fragment).	
GN	ADAMTS-11.	
OS	Oryctolagus cuniculus (Rabbit).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.	
OX	NCBI_taxid=9986;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Goad D.L., Goad M.E.;	
RT	"Molecular cloning of lapine aggrecanase-2 (ADAMTS-11) from articular	
RT	chondrocytes."	
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF317415; AAG33062.1; -.	
DR	HSP; Q9PW35; 1BDU.	
DR	MROPS; M12.425; -.	
DR	InterPro; IPR001590; Reprolysin.	
DR	InterPro; IPR000884; TSP1.	
DR	InterPro; IPR000130; Zn_MTpeptidse.	
DR	Pfam; PF01421; Reprolysin; 1.	
DR	Pfam; PF00090; Tsp.1; 1.	
DR	SMART; SM00209; TSP1; 1.	
DR	PROSITE; PS0215; ADAM_MEPRO; 1.	
DR	PROSITE; PS0092; TSP1; 1.	
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.	
FT	NON_TER 1	
FT	NON_TER 269	
SQ	SEQUENCE 269 AA; 29193 MW; 97AlCA80B33452FA CRC64;	
Query Match	20.5%; Score 830; DB 6; Length 269;	
Best Local Similarity	54.7%; Pred. No. 3.le-68;	
Matches 146; Conservative 37; Mismatches 74; Indels 10; Gaps		

```

.QY 117 TLGMADVGTMCDPKRSVIEDDGLPSAETTAHELGHVFNMPHDNVKVCCEEVFGKLRANH 176
Db 1 TLGMADVGTICSPERSCAVIEDDGLHAFTVAHEIGHLGLSHDDSKFCEENFGSTEDKR 60
QY 177 MMSPTLIQIDRANPWSACSAAIITDFLDSGHGDCLLDOPSKPISLPEDLPGASYTLSOQC 236
Db 61 LMSSILTSIDASKPWSKCTSATITEFLDDGHGNCLLDVPRKOILGPEELPGQTYDATOQC 120
QY 237 ELAFVGSKPCPYMYCTKLWCTGKAKGQMVQOTRHFPWADGTSCGEGKLCILKGACVER- 295
Db 121 NLTFGPEYTVCPGMDVCARLWCAVVRQGMVCLTKKLPAVECTPCGKGRICLQKCVDKT 180
QY 296 ----HNLNKHRYDGSWAKWDYPGCSRTCGGQVQLARRQCTNPPTPANGGKYCEGVVRKYR 351
Db 181 KKKYISTSSH---GNWGSWGPWGQCSRGCGGVQFAYRHNNPAPRNSGRYCTGKRATYR 237
QY 352 SCNLEPCPSSASGKSFREOCEAFNGY 378
Db 238 SCSVTPCP--ANGKSFREOCEAKNGY 262

```

Search completed: April 29, 2003, 17:18:02
Job time : 54.0027 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: April 29, 2003, 17:10:24 ; Search time 9.08922 Seconds
(without alignments)
3367.676 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_950

Perfect score: 4043

Sequence: 1 FVSIPIRYVETLVVADESMVK.....DQCNLHRKPQLDFCVLRPC 738

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2257	55.8	968	1	ATSL_MOUSE
2	2254	55.8	967	1	ATSL_HUMAN
3	2251	55.7	967	1	ATSL_RAT
4	1911.5	47.3	905	1	ATSL_MOUSE
5	1900	47.0	890	1	ATSL_HUMAN
6	1779	44.0	1629	1	ATSL_HUMAN
7	1761	43.6	837	1	ATSL_HUMAN
8	1745	43.2	930	1	ATSL_HUMAN
9	1730.5	42.8	630	1	ATSL_RAT
10	1728	42.7	930	1	ATSL_MOUSE
11	1184.5	29.3	1593	1	ATSL_HUMAN
12	1155.5	28.6	1077	1	ATSL_HUMAN
13	1094	27.1	997	1	ATSL_HUMAN
14	1075	26.6	1211	1	ATSL_HUMAN
15	1062.5	26.3	1205	1	ATSL_HUMAN
16	1045.5	25.9	1205	1	ATSL_BOVIN
17	942.5	23.3	860	1	ATSL_HUMAN
18	601.5	14.9	245	1	ATSL_BOVIN
19	599	14.8	207	1	ATSL_BOVIN
20	381.5	9.4	450	1	ATSL_MOUSE
21	267.5	6.6	956	1	AD19_HUMAN
22	260	6.4	776	1	AD28_MACFA
23	258.5	6.4	571	1	DISU_BOTJA
24	255.5	6.3	1584	1	BAIL_HUMAN
25	253.5	6.3	1074	1	SMSA_HUMAN
26	253	6.3	824	1	AD08_HUMAN
27	250.5	6.2	775	1	AD28_HUMAN
28	248.5	6.1	1170	1	TSP1_HUMAN
29	248	6.1	1077	1	SMSA_MOUSE
30	247.5	6.1	1093	1	SMSB_MOUSE
31	245.5	6.1	1170	1	TSP1_MOUSE
32	244.5	6.0	920	1	AD19_MOUSE
33	244.5	6.0	1170	1	TSP1_BOVIN
					P97857 mus musculus
					Q9uh18 homo sapien
					Q9wug1 rattus norv
					P57110 mus musculus
					Q9up79 homo sapien
					Q9p2n4 homo sapien
					O75173 homo sapien
					Q9una0 homo sapien
					Q9esp7 rattus norv
					Q9r001 mus musculus
					P58397 homo sapien
					Q9h324 homo sapien
					Q9ukp4 homo sapien
					O95450 h adamts-2
					O15072 homo sapien
					P79331 b adamts-2
					Q9ukp5 homo sapien
					Q9tt93 bos taurus
					P9tt92 bos taurus
					P58459 mus musculus
					Q9h013 homo sapien
					Q9xsl6 macaca fasc
					P30431 bothrops ja
					O14514 homo sapien
					Q13591 homo sapien
					P78325 homo sapien
					Q9ukg2 homo sapien
					P07996 homo sapien
					Q62217 mus musculus
					Q60519 mus musculus
					P35441 mus musculus
					Q35674 mus musculus
					Q28178 bos taurus

34	238	5.9	774	1	AD28_MOUSE
35	237.5	5.9	813	1	AD33_HUMAN
36	236.5	5.8	909	1	AD12_HUMAN
37	233.5	5.8	903	1	AD12_MOUSE
38	221	5.5	857	1	AD22_MOUSE
39	220.5	5.5	906	1	AD22_HUMAN
40	220	5.4	416	1	HR1B-TRIFL
41	219	5.4	814	1	AD15_HUMAN
42	219	5.4	826	1	AD08_MOUSE
43	208.5	5.2	1172	1	TSP2_MOUSE
44	208.5	5.2	1173	1	TSP1_XENLA
45	207.5	5.1	797	1	AD33_MOUSE

ALIGNMENTS

RESULT 1
ID ATSL_MOUSE STANDARD; PRT; 968 AA.
AC P97857; OS4768;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (SC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=98110583; PubMed=9441751;
RA Kuno K., Lizasa H., Ohno S., Matsushima K.;
RT "The exon/intron organization and chromosomal mapping of the mouse
RT ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";
RL Genomics 46:466-471(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97150761; PubMed=8995297;
RA Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,
RA Matsushima K.;
RT "Molecular cloning of a gene encoding a new type of metalloproteinase-
RT disintegrin family protein with thrombospondin motifs as an
RT inflammation associated gene.";
RL J. Biol. Chem. 272:556-562(1997).
RN [3]
RP CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.
RX MEDLINE=99303657; PubMed=10373500;
RA Kuno K., Terashima Y., Matsushima K.;
RT "ADAMTS-1 is an active metalloproteinase associated with the
RT extracellular matrix.";
RL J. Biol. Chem. 274:18821-18826(1999).
RN [4]
RP FUNCTION.
RX MEDLINE=20389568; PubMed=10930576;
RA Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,
RA Ohno H., Matsushima K.;
RT "ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";
RL FEBS Lett. 478:241-245(2000).
RN [5]
RP FUNCTION, AND INDUCTION.
RX MEDLINE=20243757; PubMed=10781075;
RA Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,
RA Richards J.S.;
RT "Progesterone-regulated genes in the ovulation process: ADAMTS-1 and
RT cathepsin L proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH

Q9jln6	mus musculus
Q9bz11	homo sapien
O43184	homo sapien
Q61824	mus musculus
Q9riv6	mus musculus
Q9p0k1	homo sapien
P20164	trimeresuru
Q13444	homo sapien
Q05910	mus musculus
Q03350	mus musculus
P35448	xenopus lae
Q923w9	mus musculus

CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-1-LEU-1692
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX.
CC
CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC CELLS OF PREOVULATORY FOLLICLES.
CC
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC
CC -1- PTM: THE PRECURSOR TO PEPTIDASE FAMILY M12B.
CC
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 7.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AB001735; BAA24501.1; ALT_INIT.
CC EMBL; D67076; BAA11088.1; ALT_FRAME.
CC
CC MEROPS: M12.222;
CC
CC MGD; MGI:109249; Admts1.
CC
CC InterPro; IPR001762; Disintegrin.
CC
CC InterPro; IPR02870; Pep_M12B_propep.
CC
CC InterPro; IPR001590; Reprolysin.
CC
CC InterPro; IPR000884; TSP1.
CC
CC InterPro; IPR000130; Zn_Mtpeptdse.
CC
CC Pfam; PF00090; tsp.1; 3.
CC
CC Pfam; PF01421; Reprolysin; 1.
CC
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC
CC SMART; SM00209; TSP1; 3.
CC
CC PROSITE; PS00215; ADAM_MEPRO; 1.
CC
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC
CC PROSITE; PS00092; TSP1; 3.
CC
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC
CC Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix; Heparin-binding.
CC
CC SIGNAL 1 48
CC PROPEP 49 253
CC CHAIN 254 968
CC SITE 206 206
CC METAL 402 402
CC ACT_SITE 403 403
CC METAL 406 406
CC METAL 412 412
CC DOMAIN 477 559
CC DOMAIN 560 617
CC DOMAIN 618 725
CC DOMAIN 726 850
CC DOMAIN 851 909
CC DOMAIN 910 968
CC DOMAIN 195 199
CC CARBOHYD 548 748
CC CARBOHYD 721 721
CC CARBOHYD 765 765
CC CARBOHYD 783 783
CC CARBOHYD 946 946
CC MUTAGEN 403 403
CC CONFLICT 335 335
CC CONFLICT 425 425
CC SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;

Query Match 55.8%; Score 2257; DB 1; Length 968;
Best Local Similarity 53.6%; Pred. No. 1.2e-156;
Matches 400; Conservative 130; Mismatches 176; Indels 40; Gaps 10;

QY 1 FVSPRVVETLVVADESVMVKFHGADLEHYLLTLLATAARLYRHPSILNPNIVVVKVLL 60
DB 254 FVSPRVVETMLVADQSWADFHGSLKHYLLTFLSVARFYKHPKSIRNSISLVVVKILVI 313
QY 61 RDRSQPVKTGNAALTLENFCWQKLNKVKSDHPEYDWTAILFTRODLGCGTCDLGM 120
DB 314 YEEQKPEVTSNAALTLENFCWQKQNSPDRPEYDWTAILFTRODLGCGTCDLGM 373
QY 121 ADVGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180
DB 374 ADVGTVCDPSRSCSVIEDDGLQAFTTAHELGHVFNPHDDAKHCASLVGTGSHLMAS 433
QY 181 TLTIQIDRANPWSACSAAITITDLSGHGCDLLDOPSKPISLPEDLPGLGASYTSLQOCELAF 240
DB 434 MLSSLDHSQPSPCSAVYVTFSLDNGHGECLMDKPNKIPKLPDLPGTLYDANRCQFTF 493
QY 241 GYSGKPCP-YMOYCTKLMCTGKAKGQMVCOYTRHFPWADTSCGEGKCLCKACACVERHNLN 299
DB 494 GEESKHCFDAASTCTTLCWCTGTSGLLVCQTKHFPWADTSCGEGKVCVKYKNTDM- 552
QY 300 KH---RVDSWAKWDPYGPCSRTCGGVLARRCTNPTANGGKYCRGVKVRYSNCLE 356
DB 553 KHFPATVHSGMGPWGPWGDSCSTCGGVQYTMRECDNVPKNGGKYCRGVKVRYSNCIE 612
QY 357 PCPSASCKSPREECEAFNGYNHSTNRLTLAVAMVPKYSGVSPRDKCKLICRANGTYGF 416
DB 613 DCPDN-NKGTFFREEQCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKLTCEAKGIGYF 671
QY 417 YVLAPKVVDTLCSPTDSTSVCGQKICRAGCDNGLSKRRFDKCGVCGDNKSKKVTGL 476
DB 672 FVLQPKVVDTGTPCSPDSTSVCGQCVKAGCDRIIDSKKFKDCGVCNGSTCKKMSGI 731
QY 477 FTKPMHGYNFVAIPAGASSIDIRORYKGLIGDDNYLALNKSQKYLNLGHFVVSAYER 536
DB 732 VTSTRPGYHDVITIPAGATNIEVKHNRONGSRNNGSFALRAADGTIYILNGFTLSTLEQ 791
QY 537 DLVWKGSLRYSGTGTAVESLQASRPILPLEVLSVSGKMTPPRVRYSYFLPKPRBDK 596
DB 792 DLTGKTVLRYSSSSAALERISFSPLEPLTIQVLWGHALRPKIKFTYFMKKKTES-- 849
QY 597 SHPKDPRGSPVHLNLSVLSNQVEQDDRPAPRVAGSWGSCSAGSLQKRAVDCRG 656
DB 850 -----FNAIPTFS-----EWVIEWEGCSKTCGSGWRRVVQCRD 884
QY 657 SAGORTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGFQRRSLKCVGHGG 713
DB 885 INGH---PASECAKEVKVPASTRPCADLPCHWQVGDWSPCSKTCGKYKRTLKCVSHDG 941
QY 714 RLARDQCNLHRRKQPE-LDFCVLRPC 738
DB 942 GVLNSCSDPLKKPKHYIDFCTLTCQ 967

RESULT 2
ID ATSL_HUMAN STANDARD; PRT: 967 AA.
AC Q9UHT8; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
GN ADAMTS1 OR METH1 OR KIAA1346.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.

Db 215 FVSEARFVETLTVADASMAAFYAGDQLQNHILTLMSVAARIYKHPISIKNSINLMVVKVLIV 274
QY 61 RORDSGPKVTGNAALTNRNFCWQKLNKYSKDHPEYWDTAILFRQDLGAT-TCDFILG 119
Db 275 EBEKVGPEVSDNGGLTLNFRNFWQRFRNQPDSRHPHEHYDTAILLFRQFCGQEGCLDTLG 334
QY 120 MADVGTMCDPKRSKSVIEDGLPSAFTTAHELGHVFNPHDNVNVKCEVEFGKLRANHMS 179
Db 335 VADIGTICDPNPKSCSVIEDGELQAHTLAHELGHVLSLPHDSDCKPCTRLFPGMKHHVNA 394
QY 180 PTLIQIDRANPASCASAIITDFLSGHGDCLLDQSPKISLPEDLPG--ASYTILSQOCE 237
Db 395 PLFVHLNQTLPWSPCSAMYLTELLDGHGDCLLDAPGAALPLTGLPGRMALYQLDQOCR 454
QY 238 LAFGVSKPCP---YMQVCTKLWCTGKAGQWVCOTRH--FPWADGTSCGEGKCLKAGC 292
Db 455 QIFGPDFRCHPNTSADQVCAQWCHTDG-AEPLCHTKNGSLPWADGTPCGPHLCSEGC 513
QY 293 VERHNLNKHRR--VDGSAKWDPYGCRTCCGGVQLARRQCTNPTPANGKYGKCEGVVKY 350
Db 514 LPDEEVEVRPKPVVDGWPAGWPGWBCSRTCCGGVQFSHRECKDPEPQNGGYCLGRRAKY 573
QY 351 RSCNLEPCPSSASGHSFEEQCEAFNGYHNSTNRLTLAVAVPKYSGVSPRDKCKLICRA 410
Db 574 QSCHETECP--PDGKSFREOQCEKYNNAYTMDGNL-LQWVPKYAGVSPRDRCKLFCRA 630
QY 411 NGTGYFYVLAPKVVDTGLCSPDSTSVCGVKICRAGCDGNLKKRFDKCVCGDKNKSC 470
Db 631 RGRSEKFEAKVIDGTLGCPETALCVRGOCVRAGCDHVVDSPRKLDDKCGVCGGKNSC 690
QY 471 KVTGLGTFKPMHGVNVVAIPAGASSIDIRQYKGLIGDDNYLALKNKSGKYLINGHFV 530
Db 691 RYVSGSLPTNYGNDIVTIPAGATNIDVKORSHPGVQNDGNLYLALKTADGQYLLNGNLA 750
QY 531 VSAVERDLVWKSLLRYSCTGAVESLOASRPILPLETLVEVLSV-GKMTPPRVRYSYLP 589
Db 751 ISAIQDILVGLTKLYSGSIATLRLQSFRLPELPLTVQLLTPVGEVPEKRYTFFVP 810
QY 590 KEPRDKSHPKDRPGSVLHNSVLSLSNQVEQPDPRPARWAGSWGPCSASCGSLQK 649
Db 811 ND--VDFSMQSKERATNIIQLLH-----AQWVLGDSECSSTCGACWQR 855
QY 650 RAVDCRGAGORTVPACDAARHPVETQAC-GEPCP 683
Db 856 RTVECRDPGQASA-TCNKALKPEDAKPCESQLCP 889
RESULT 6
ID AT59_HUMAN STANDARD; PRT; 1629 AA.
AC Q9P2N4; Q9NR29;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
GN ADAMTS9 OR KIAA1312.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Fetal;
RX MEDLINE=20396138; PubMed=10936055;
RA Clark M.E., Kellner G.S., Turbeville L.A., Boyer A., Arden K.A.,
RA Maki R.A.;
RT "ADAMTS 9, a novel member of the ADAM-TS/Metallospondin gene
PL family",
PL Genomics 67:343-350(2000).
RP [2]
RN SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;

RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RL for large proteins in vitro";
RL DNA Res. 7:65-73(2000).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.
CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,
CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN
CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR
CC THYMUS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF261918; AAF89106.1; -;
DR EMBL; AB037733; BAA92550.1; -;
DR HSSP; P15167; IATL.
DR MEROPS; M12.021; -;
DR Genew; HGNC:13202; ADAMTS9.
DR MTM; 605421; -;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF00090; tsp_1; 11.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPl; 12.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS50092; TSPl; 9.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 18
FT PROPEP 19 287
FT CHAIN 288 1629
FT DOMAIN 509 587
FT DOMAIN 589 642
FT DOMAIN 645 752
FT DOMAIN 753 880
FT DOMAIN 999 1053
FT DOMAIN 1056 1108
FT DOMAIN 1111 1156
FT DOMAIN 1184 1239
FT DOMAIN 1240 1295
FT DOMAIN 1332 1383
FT DOMAIN 1386 1439
FT DOMAIN 1445 1498
FT DOMAIN 1501 1554
FT DOMAIN 1562 1612
FT DOMAIN 88 96
FT SITE 223 223
CYSTEINE SWITCH (POTENTIAL).

FT	METAL	434	434	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	435	435	BY SIMILARITY.
FT	METAL	438	438	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	444	444	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	112	112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	135	135	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	271	271	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	749	749	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	840	840	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1213	1213	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1267	1267	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1064	1072	CLVTCGKGH -> VRWEGCYFP (IN SHORT ISOFORM).
FT	VARSPLIC	1073	1629	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	367	367	F -> L (IN REF. 1).
FT	SEQUENCE	1629	AA; 182649	MM; C1C4CEFF58B8941F CRC64;
Query Match				
Best Local Similarity		44.08;	Score 1779;	DB 1; Length 1629;
Matches 354;		Conservative 115;	Mismatches 250;	Indels 66; Gaps 16;
Qy	1	FVSPRYVETLVADSMVKFADLEHLYLTLLATLAARLYRHPSILNPINIVVKKVLL	60	
Db	288	FLSPRFVEVLVDNRVSTHGENLOHYILTMSIVASYKDSIGNLNINIVNLVI	347	
Qy	61	RDRSGPKVTGNAALTLRNCAWQKLNKYSKHEPYWDTAILFTRODLCA-TTCDTLG	119	
Db	348	HNEQDPSISFNAOTTLKNFCQWQHSNPGGIH---HDTAVLLTRQDICRAHDKCDTLG	404	
Qy	120	MADVGTMCDPKRSVSIEDDGLPSAFTAHGLGHVFNPHDNVKNVCEVFGKLRANHMS	179	
Db	405	LAELGTICDPYRSCSISDSGLSTAFTAHGLGHVFNPHDNVKNVCEVFGKLRANHMS	463	
Qy	180	PTLIQIDRANPWSACSAIITFDLSHGDCLLDQP-SKPTSLPEDLPASGYTSLQOCEL	238	
Db	464	PTLNFYTPNWMKSCRXYITFDLTGEGCLLNEPESRPPYLPVQLPGILYNNKQCEL	523	
Qy	239	AFGVGSKPCPNQYQYCTKLWC---TGKAKQMVQCTRIHPFWDGTCGEGKLCAGCVER	295	
Db	524	IFPGSQVCPYQMVQCRRLWCNNVGVHKG---CRTQHTPWADGTCEPBGKHCYGFVCPK	580	
Qy	296	HNLNKRVGDSWAKWDPGPCSRCTCGGVQVLARQCTNPTPANGKYCEGVVRYKSCNL	355	
Db	581	EMDVPTVDGSMGWSPEGTCSRCTCGGVQVLARQCTNPTPANGKYCEGVVRYKSCNL	639	
Qy	356	EPCPSASGKSPREQCEAFNGYHSTNRLTLAVAWPKYSGVSPROCKKLCIRANGTY	415	
Db	640	EPCLAKQ---RDFRQCAHFDGKHENINGLLPNVRWPKYSSILMKDCKLFCRVAGNTA	697	
Qy	416	FVVLAPKVVDTGLSPDSTSCVQKGLKAGCDGNLMSKRPDKCGVCGGDNKSKKVTG	475	
Db	698	YYQLRDRVIDGTPCGQDNDICVQGLCRQAGCDHYLNSKARRDKCGVCGGDNSSCKTVAG	757	
Qy	476	LFTKPMHGNFVWATPAGASSITDRQYKGLIGDNDYLNKNSQGYLLNGHFVWSAVE	535	
Db	758	TFTVHYGTIVTVRIPAGATNIDVRQHSGETDDNDYLNKNSQGYLLNGHFVWSAVE	817	
Qy	536	RLVVKGLSLRYSGTGTAVESLQASRIEPLTVEVLSVGMKTPPRVRYSLPKPEPRED	595	
Db	818	REIRIGNAVVEYSGSETAVERINSTDRIEQELLQVLSVGLKYNPDVRYSNIPEDKPQ	877	
Qy	596	K---SSHPRKDRPGP-----SVLNSVLSLSNQVEQPDPRPP-----	628	
Db	878	QYWNH-----GPMQACSKPCQGERKRLVCTRESQDLTVS---DQRCDRLPQFGHITE	929	
Qy	629	-----ARWVAGSWGPCASGGLQKRAVDCR-----GSAGORTVPACDAARHPVT	675	
Db	930	PGCTODDLRWVAVSSESAQGLGYRDLTYCAKYSRLDGTETKVDGFCSSHPKPSNR	989	
Qy	676	QACGEPCTP---WELSAWSPCKSCGGRFQRRSLKVCVGHGRLARDQCNLHRKQFQELDFC	733	
Db	990	EKCSGECNTGWRYSATWCKSCDGGTORRRAICVNTNRDVLDDSKCT-HQEKVTIQRC	1048	

QY	734	VLRPC	738	
Db	1049	SEFFC	1053	
RESULT 7				
IDS	ATSA_HUMAN	STANDARD;	PRT;	837 AA.
AC	O75173; Q9UN83;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ADAMTS-4 precursor (fc 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1) (ADMP-1).			
GN	ADAMTS4 OR KIAA0688.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=98403880; PubMed=9734811;			
RA	Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,			
RA	Kotani H., Nomura N., Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes. X.			
RT	"The complete sequences of 100 new cDNA clones from brain which can			
RT	code for large proteins in vitro.";			
RL	DNA Res. 5:169-176(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99286303; PubMed=10356395;			
RA	Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,			
RA	Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,			
RA	Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,			
RA	Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,			
RA	Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,			
RA	Trzaskos J.M., Arner E.C.;			
RT	"Purification and cloning of aggrecanase-1: a member of the ADAMTS			
RT	family of proteins.";			
RL	Science 284:1664-1666(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;			
RT	"ADAMTS-4 genomic locus.";			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	PARTIAL SEQUENCE, AND CHARACTERIZATION.			
RX	MEDLINE=20400518; PubMed=10827174;			
RA	Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,			
RA	Burn T.C., Arner E.C.;			
RT	"The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for			
RT	aggrecan substrate recognition and cleavage.";			
RL	J. Biol. Chem. 275:25791-25797(2000).			
CC	- - FUNCTION: CLEAVES AGGECAN. A CARTILAGE PROTEOGLYCAN, AND MAY BE			
CC	INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE			
CC	DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE			
CC	CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN			
CC	ALZHEIMER'S DISEASE.			
CC	- - CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu- -Ala-393			
CC	site.			
CC	- - COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	- - SUBCELLULAR LOCATION: Secreted. Associated with the extracellular			
CC	matrix (by similarity).			
CC	- - TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED			
CC	AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.			
CC	- - INDUCTION: BY INTERLEUKIN-1.			
CC	- - DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT			
CC	FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.			
CC	- - PTM: THE PRECURSOR IS CLEAVED BY A FUFIN ENDOPEPTIDASE.			
CC	- - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.			
CC	- - SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.			

RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
zinc metalloproteases.";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-1-Ala-393
site.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (by similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,
CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
ARTHRITIC PATIENT.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF142099; AAD49577.1; -
CC EMBL; AP001698; BAA95504.1; -
CC EMBL; AP001697; BAA95503.1; -
CC EMBL; AF141293; AAF02493.1; -
CC HSP; Q9PW35; 1BUD.
CC MEROPS: M12.225; -
CC Genew; HGNC:221; ADAMTS5.
CC MIM; 605007; -
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSPL.
CC InterPro; IPR000130; Zn_M12Bptdse.
CC Pfam; PF00090; tsp_1; 2.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSPL; 2.
CC PROSITE; PS0215; ADAM_MEROPS; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS00092; TSPL; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 16
FT PROPEP 17 261 POTENTIAL.
FT CHAIN 262 930 ADAMTS-5.
FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 411 411 BY SIMILARITY.
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 566 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE-1 1.
FT DOMAIN 624 731 CIS-RICH.
FT DOMAIN 732 874 SPACER.
FT DOMAIN 875 930 TSP TYPE-1 2.
FT DOMAIN 37 41 POLY-ALA.
FT DOMAIN 257 261 POLY-ARG.

FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 138 138 A -> G (IN REF. 2).
FT CONFLICT 614 614 R -> H (IN REF. 3).
FT CONFLICT 692 692 P -> L (IN REF. 2).
SQ SEQUENCE 930 AA; 101715 MW; B64281502F28193B CRC64;

Query Match 43.2%; Score 1745; DB 1; Length 930;
Best Local Similarity 48.7%; Pred. No. 2.le-119;
Matches 335; Conservative 98; Mismatches 219; Indels 36; Gaps 10;

QY 2 VSPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHSPTLNINIVKVLRLR 61
Db 263 ISRAQVELLVADASMARLYGRGLQHYLLTLASTANRLYSHASTENIRLAVKVVVLG 322
QY 62 DRDGPYKVTGNAAFLRNFCANQKLNKVDKHPYWTAILFTRODLCGATTCDTGLMA 121
Db 323 DKDKSLEYSKNAATLKNFCKWQHQNLDGDDHEHYDAAILFTREDJGHHSCDTLGM 382
QY 122 DVTGWCDPKRSCSVTEDDGLPSAFTTAHELGHVFNPHDVKVCEEVFKLRANHMSPT 181
Db 383 DVTGTCSPERSCAVIEDDLGHAFTVAHEIGHLGLSHDDSKFCEETFGSTEDKRLSSI 442
QY 182 LIQIDRANPWSACSAAITDFLDSHGDCCLLDQPSKPSLSPEDLPASVYTLSSQCELAFG 241
Db 443 LTSIDASKPWSKTSATITEFDLDGHNCLDLPRKQLGLPEELPGQYDATQCNLTFG 502
QY 242 VSGKPCPYMYCTKWCCTKAGQWVCPTRHPFWDGSCGKCLKLGACVER----H 296
Db 503 PEYSVCPGMDVCAWLCAVVRQGMVCLTKLPVAGEGTPCGKGRICLOGKCVDTKKYY 562
QY 297 NLNKHVRDGSWAKWDPYGPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVRYKYSNCL 356
Db 563 STSSH---GNWGSWSGSGCSRSCGGVQFAYRHCNPNAPRNGRYCTGKRAIYRCSILM 619
QY 357 PPSSASGKSFRECEQAFNGYNHSTNRLTLAVANVPKSYGSPDKCKLICRANGTVGF 416
Db 620 PCP--PNGKSFRECEQAKNGYQSDAKGVKTFVWVPKYAGVLPADVCKLTCAKGTGY 677
QY 417 YVLAKPVVDGTLTCLSPDSTSVCGQKCIKAGCDNGLSKKRFKDCGVCGGDNKSKKVTGL 476
Db 678 VVFPKVTGTECPYNSVYRGCVRGCDIIGSKLQYDKGVCGGDNSSCTKIYGT 737
QY 477 FTKPMHGYNEVVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQKYLNLGHFVYSAVER 536
Db 738 FNKKSQYTDVVRIPGEGATHIKVQPKAKDQTRFTAYLALKKKNGEYLINGYIMISTSET 797
QY 537 DLVVGSLRLRYSGTGTAVESL-----QASRPILPLETVSVLSVGMKMTPPRVYSFVLPKE 591
Db 798 IIDINGTVMNSYSGWSHRDDFLHGMGYSAIKEL---LIVQILATDTPKPLDVRYSEFVPK 854
QY 592 PREDKSHPKDPRGSPVLSHNSVLSL-SNQVEQPDPRPPARVWAGSWGPCASCGSLQKR 650
Db 855 -----STPK-----VNSVTSHGSKNKGSHTSQP--QWVTGPMWLCASRTCDTGWHTR 898
QY 651 AVDCRSAGQRTVPACDAHRPVETOAC 678
Db 899 TVQCQ-DGNRKLAKGCPLSQRPSPAFKQC 925

RESULT 9
ID AT54_RAT STANDARD; PRT; 630 AA.
AC Q9ESP7; Q9ESP8; Q9ESP6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (Fragment).
GN ADAMTS4.

QY	122	DVGTCDCEKRCSEVIEDGLPSAFTTAHELGHVFNPNHONKVCEVEFGVKLRANHMMSPT	181
Db	383	DVGTTCSPERSCAVLEDDGLHAATVAHEIGLHLGLSHDSDKFCEENFCTTDEKRLSSI	442
QY	182	LFIQIDRANPWSACSAIITDFLDGSHGDCLLDOPSPISLPEDLPASVTLSOQCELAFG	241
Db	443	LTSIDASKPWSKSTSATITEFLPDGHGNCLLDLPRKQILGLPELPGQTYDAYQQCNLTFFG	502
QY	242	VGSKPCPYMQVYCTLKWCTGAKAGOMVCOTRHFPWADGTCGEGLCKLGACAYER----	296
Db	503	PYSYCPCGMDCVARUWCVAWFQGMVCLTKLPAVEGTFCGKGRCVLQGCVDKTKKKYY	562
QY	297	NLNKHVRVDSNAKWMDPYGPCSRTCGGVQLARRQCTNPTPANGKYKEGVSRDKCLICRANGTVGF	356
Db	563	STSSH---GNMGSGCPWGQCSRSRSCGGVGQFAYRHCNNPAPRNSGRYCTCKRAIYRSCSVT	619
QY	357	PCPSSASGKSFREDOCAFNGYNHTRLLAVANVPKYSGVSPPROCKLICRANGTVGF	416
Db	620	PCP--PNGKSPRHEOCSEAKNGQDAKGVKTFEVMFKYAGVLPADVCKLCRAKGTGY	677
QY	417	XVLAPKRVDTLCSPDSTSVCVOGCKTKAGCDGNLGSKKRFDCGCVGGDNKSCRKVTGL	476
Db	678	VTFSPKVDGTETCRYPYNSVCVRGECVRTGCDGLIGSLKLOYDKCGVGGDNSCTKIIGT	737
QY	477	FYKPMHGYNFVVAIIPAGASSIIDIRQRYKGLIGDDNYLALKNSOGXYLLNGHVVSAYER	536
Db	738	FNKSKSGYTDVVRIPEGATHIKVQFRKAQDQTPPAYLALKKKTGETYLINGKYMISTSET	797
QY	537	DLVVGSLRLRYSGTGTAYESL-----QASRPILLEPLNEVLNVSGKMTPTTPRVYSFLPKR	591
Db	798	IIDINGTVMNYSGWHRDDFLHGMGYSATKEI----LIVQILATDPFKALGVRISSFVPKK	854
QY	592	PREDKS---SHPKDPRGSPVLSHNVSLSLSNQVEQDDRRPARVWAGSWGSPCASGSGLO	648
Db	855	TTOQVNSVISHSNKVGPI--HSYQL-----QWVTEPWLACSTRCTDGTWH	896
QY	649	KRAVDCRGSGAQRTVPACDAAHPRVETOAC	678
Db	897	TRTVQCO-DGNKRKLAKGCLLSORPSAFQOC	925
RESULT 11			
AT12_HUMAN			
ID	AT12_HUMAN	STANDARD;	PRT; 1593 AA.
AC	P58397;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	ADAM-TS-12 precursor (EC 3.4.24.-) (A disintegrin and metallopeptidase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12).		
GN	ADAMTS12.		
OS	Eukaryota (Human).		
OC	Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Fetal lung;		
RX	MEDLINE=21264577; PubMed=11279086;		
RA	Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;		
RT	"Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin-1 repeats.";		
RL	J. Biol. Chem. 276:17932-17940(2001).		
CC	-!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).		
CC	-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).		
CC	-!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely expressed in gastric carcinomas and in cancer cells of diverse origin.		
CC	-!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.(BY		

CC CC SIMILARITY).

CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.

CC -1- PTM: IS SUBJECT TO AN INTRACELLULAR MATURATION PROCESS LEADING

CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE

CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND

CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1

CC DOMAINS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AJ250725; CAC20419.1; -

CC Genew; HGNC:14605; ADAMTS12.

CC MIM; 606184; -

CC InterPro; IPR001762; Disintegrin.

CC InterPro; IPR002870; Pep_M12B_propep.

CC InterPro; IPR001590; Repolysin.

CC InterPro; IPR000884; TSP1.

CC InterPro; IPR000130; zn_Mtpeptdse.

CC Pfam; PF00090; tsp-1; 6.

CC Pfam; PF01421; Repolysin; 1.

CC Pfam; PF01562; Pep_M12B_propep; 1.

CC SMART; SM00209; TSP1; 8.

CC PROSITE; PS0215; ADAM_MEPRO; 1.

CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.

CC PROSITE; PS0092; TSP1; 2.

CC PROSITE; PS0042; ZINC_PROTEASE; 1.

CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

CC Repeat; Extracellular matrix.

CC

CC SIGNAL 1 25 POTENTIAL.

CC FT PROPEP 26 240 BY SIMILARITY.

CC FT CHAIN 241 1593 ADAMTS-12.

CC FT DOMAIN 465 544 DISINTEGRIN-LIKE.

CC FT DOMAIN 545 596 TSP TYPE-1 1.

CC FT DOMAIN 597 700 CYS-RICH.

CC FT DOMAIN 701 826 SPACER 1.

CC FT DOMAIN 827 881 TSP TYPE-1 2.

CC FT DOMAIN 886 943 TSP TYPE-1 3.

CC FT DOMAIN 947 995 TSP TYPE-1 4.

CC FT DOMAIN 996 1315 SPACER 2.

CC FT DOMAIN 1316 1364 TSP TYPE-1 5.

CC FT DOMAIN 1367 1423 TSP TYPE-1 6.

CC FT DOMAIN 1426 1471 TSP TYPE-1 7.

CC FT DOMAIN 1426 1471 TSP TYPE-1 8.

CC FT DOMAIN 302 305 POLY-GLU.

CC FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).

CC FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).

CC FT ACT_SITE 393 393 BY SIMILARITY.

CC FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).

CC FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).

CC FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 685 685 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 790 790 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. .) (POTENTIAL).

CC FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. .) (POTENTIAL).

CC SEQUENCE 1593 AA; 177545 MW; 07F9F48563BD83A3 CRC64;

Query Match 29.3%; Score 1184.5; DB 1; Length 1593;

Best Local Similarity 34.4%; Pred. No. 2.6e-78;

Matches 279; Conservative 122; Mismatches 280; Indels 129; Gaps 29;

QY 2 VSIPRYVELTVADSMVKFHGAD-LEHYLLTLTATAARLYRHPHSILNPNVVKVLL 60

DB 242 ISKERWVETLVVADTKMIEYHGSENVESYILTIMNWTGLFHNPSIGNAIHIVVRILL 301

QY 61 RORDSGPKVTGNAALTLRNFCWQKLNKVSQDKHPEYWDTAILETRDOLCGA--TTCDTL 118

DB 302 EEEQGLKIVHAEKTLSSFCWKQKSIKPKSOLNPHVDVAVLLTRKIDICAGENRCPETL 361

QY 119 GNADVTGMDCKRSCSVIEDDGLPSAFTAHGLHGVFNMHPDNVK-VCEEVFGKLRANHM 177

DB 362 GLSHSGMCOPIHRSNINSDSLPLAFTAHGLHSGFIQHDGKENDCEPVG--RHPIYI 418

QY 178 MSPTLTIDRANPWSACSAIIITDFLDSDGHGDCILDDQPSKP-ISLPEDLPGASYTL 236

DB 419 MSRLQYDPTPLTWSKSEYITRELDRGWGFCFLDDIPKKGLKSKVIAPGVYDVHHC 478

QY 237 ELAFGVSRPCPYMQ-YCTKLWCTGAKQMVQOTRHPFWADGTSCGEGKLCGLGACVER 295

DB 479 QLOYGNATFCQEVENVQCOTLWCYKG----FCRSKLDAADAGTQCGEKKWCMAGKCI-- 532

QY 296 HNLNK--HRVDGSAKWDYPCPSRTCGGVVOLARRQCTNPTPANGKYCEGVVRKYRSC 353

DB 533 -TVGKAPESIPGGWGRWSPWCHSCTCGAGVQSAERLCNNPEPFEGGYCYTGERKRYELC 591

QY 354 NLEPCPSSASGKSFREEQCAFNGYNHSTNRLTLAVAVPKYSGVSPRDKCLKICRANGT 413

DB 592 NVHPCRSEA--PTFRMQCSEFDTVPYK-NEL---YHWPPIF--NPAHPCELYCRPIDG 642

QY 414 GYFVVLAPKVVDGTLG--SPDSTSVQVQCKIKAGCDGLGSKKREDKCGVCGGDNKSK 471

DB 643 QFSEKMLDAVIDGTPCFEGGSRNVCINGICAKMVGCDTEIDSNATEDRCVGLDGGSCQ 702

QY 472 KVTGLF-TKPMHGYNFVAIPAGASSIDIRQRYKGLIGDDNLYALKNSQ-GKYLNGHF 529

DB 703 TVRKMEFKEGSGYVDIGLIPKARDIRVME-----IEGAGNFLAIRSEDEPKYYLNGF 757

QY 530 VWSAVERDLVVGKSLRYSGTGAVESLOASRPILPLTVELSVGKMTPTPRVYSFLP 589

DB 758 IIO-WNGNYKLACTVFOYDRKGD-LEKLMATGPTNESVMIQLLF--QVTNPGIRYETIQ 813

QY 590 KEPREDKSSHPKDPGRGFSVLHNSVLSLNOVEQ----- 622

DB 814 KD-----GLDNDVEQMYFWQYGHWTGECSTVCTGTRQTAHC 850

QY 623 -----PDRP-----PARWVAGSWGPCSSASCG-SGLQKRAVDICR 655

DB 851 IKKGRGMVKATFCDPETOPNGRQKKCHERACPPRWAGWEACATCGPHGKKRTVLCI 910

QY 656 GS--AGORTVPACDAAH--RPVETOACGP--CPT-WELSAWSPCKSGCGRFRRLSKC 708

DB 911 QTVWSDEQALPTDCCOHLKPKTLTLLSCNRDILCPDWTGVNWSSECSVSCGGVRIYSVC 970

QY 709 VGHGRLLRDQCNLHRKPOELDFCVLRPC 738

DB 971 AKNHD-----EPCDVTRKPNRSLALCGLQOC 995

RESULT 12

ID AT10_HUMAN STANDARD; PRT; 1077 AA.

AC Q9H324;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase

DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (fragment).

GN ADAMTS10.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
thrombospondin type 1 repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: AF163762; AAC35563.1; -;
DR MEROPS: M12.235; -;
DR Genew: HGNC:13201; ADAMTS10.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_MTpeptdse.
DR Pfam: PF00090; tsep1; 5.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR SMART: SM00209; TSP1; 5.
DR PROSITE: PS50215; ADAM_MEPRO; 1.
DR PROSITE: PS50092; TSP1; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydrolase: Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT NON_TER 1
FT PROPEP <1 207
FT CHAIN 208 1077
FT METAL 366 366
FT ACT_SITE 367 367
FT METAL 370 370
FT DOMAIN 376 376
FT DOMAIN 434 520
FT DOMAIN 578 679
FT DOMAIN 680 802
FT DOMAIN 521 577
FT DOMAIN 799 860
FT DOMAIN 862 918
FT DOMAIN 922 976
FT DOMAIN 981 1031
FT CARBOHYD 64 64
FT CARBOHYD 196 196
FT CARBOHYD 297 297
FT CARBOHYD 714 714
FT CARBOHYD 769 769
FT CARBOHYD 866 866
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCBBF587 CRC64;
Query Match 28.6%; Score 1155.5; DB 1; Length 1077;
Best Local Similarity 34.0%; Pred. No. 21e-76;
Matches 291; Conservative 101; Mismatches 304; Indels 159; Gaps 27;
QY 2 VSPRVTETLVADSWKVFHG-ADLEHYLLTLLATAARLYRHPSTILNPINIVVVKVLL 60

Db 209 VSRERYETLVADKMWAYHGRDRVEQYVLAIMNIVAKLFQDSLSGTVNLLVTRLILL 268
QY 61 RDRDSGPKVTGNAALTNRNFCWQKLL-----NKVSDKHPEYWDYTAILTRQDLG--G 111
Db 269 TEDQPTLEITHHAGSLDSFCWKQKSYNVHSHGNAIPENGVAHDTAVLTRDYDICIYK 328
QY 112 ATTCDTGLMADYGTMCDDPKRSCSIEDDGLPSAFTTAHELGHVFNHNDNVKVCVEFG- 170
Db 329 NKPCGTLGLAPVGGMCEERSCSVNEDIGLPOAFTIAHEIGHTFGMNHGVGNSCGARGQ 388
QY 171 ---KLRAHMMSEFTLIQIDRANP--WSACSAALITDFDSGHGDCLLDQPSK-PISLPED 224
Db 389 DPAKLMAAHITMKT-----NPFVWSSCNRYIITFSDSLGSLCLNNRPPRQDFVYFV 441
QY 225 LPGASYTLSCQCELAFGVGSKPCPYMAYCTKLWCTGKAKGMVCQTRHFPWADGTSCEG 284
Db 442 AFGQADADEQCFQHGKVSROCKYGEVCSLWCLSKSNR---CITNSIPAAEGTLCOYH 498
QY 285 KL-----CLGACVERINLKNHRVDGSAWAKDYPGCSRTCGGGVQLARRQCTNPTPANGG 340
Db 499 TIDKGWCYKRVCP--FGSRPEGVDGAWGPWTWGDGCSRCTGGVSSSRHCDSPRTIGG 557
QY 341 KYCEGVYRVKRCNLEPCPSSASGSKFEEQCEAFNG-----YNHSTNRLTLAVAWVP 393
Db 558 KYCIGERRRHSNCTDDCPGGS--QDFREVQCSEFDSIFRGKFKWKTYR----- 606
QY 394 KYSGVSPRDCKLICRANGTYFYVLAPKVVDTGLCSPTSTVCVQGGKCIKACGDNGLGS 453
Db 607 -GGGVK---ACSLTSLAEGFNFYTERAAAVVDGTPCRPTVDICVSGECKHVGCDRLVGS 662
QY 454 KRFDFKCVGCGGDNKCKKVTGLFT--KPMHGYNFVVAIPAGASSIDIRQGYKGLIGDD 511
Db 663 DLREDKRCVGGDGSACETIEGVSPASGAGYEDVWIPKGSVHFIOD-----LNLSL 717
QY 512 NYLAKNSQGYLLNGHFVYSAVERDLVVKSLRLYSGTGTAVESLQASRPILPTVEV 571
Db 718 SHLAKGDQESLLLEGLPGTPQPHR-LPLAGTTFQLRQGPDOVQSLEALGPINASLIVW 776
QY 572 LSVGKMTPPRVRYSYFLP----- 589
Db 777 LARTEL--PALRYRFNAPTARDLSPYSWHYAPWTKSAQAGGSOVQAVECRNQLDSSA 834
QY 590 -----KEPREDK--SSHPKDP----- 619
Db 835 VAPHYCSAHSKLPKQKQACNTEPCPDWVGVNWSLCSRCDAGVRSKSVVQRRVSAEE 894
QY 620 VEQ-----PDRPPA-----RWVAGSWGSPCSACGSLQKRAVDCRGSACQRTV 663
Db 895 KALDDSDACQPPRPVLEACHGTCPPPEWAALDWSECTPCGPGLRHVRVVLCKSADHRATL 954
QY 664 PA--CDAAHRPVTQAGC-EPGP--TWELSAWSPCSKSGRQFRRSLKCVGHGGRLLAR 718
Db 955 PPAHCSPPAKPPATMRCNLRRCPPARVWAGWEGCSAQCGVQGRQSVRCTSHTQ--AS 1012
QY 719 DQCNLHRKQELDFC 733
Db 1013 HECTEALRPPTQOC 1027
RESULT 13
AT57_HUMAN STANDARD; PRT; 997 AA.
ID Q9UKF4; 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
GN ADAMTS7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=99395124; PubMed=10464288;
Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
"ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of
Zinc Metalloproteases";
J. Biol. Chem. 274:25555-25563(1999).
CC -|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -|- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,
LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AF140675; AAC56358.1; -;
DR HSSP; P15167; IATL.
DR MEROPS; M12.231; -.
DR Genew; HGNC:223; ADAMTS7.
DR MM; 605009; -;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12b_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTpeptidse.
DR Pfam; PF00090; tsp.1; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12b_propep; 1.
DR SMART; SM00209; TSPI; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPI; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metallopeptidase; zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 538 594 DISINTEGRIN-LIKE.
FT DOMAIN 595 697 TSP TYPE-1 1.
FT DOMAIN 698 914 CYS-RICH.
FT DOMAIN 915 990 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match 27.1%; Score 1094; DB 1; Length 997;
Best local similarity 37.5%; Pred.No. 5.8e-72;
Matches 280; Conservative 90; Mismatches 301; Indels 76; Gaps 27;

QY 2 VSIPIRYVETLVVADESNVKFHG-ADLEHYLLTLTATAARYRHPSTLNIPNIVKVL 60
|| ::||||| ||::|| | :||:: | | ||| ||| :||::||

QY	561	RTVPACDAAH-----RPVETQACG-BECP-TWELSAWSPCSKSGRFFORSLKCVGHGR	714
Db	949	NTRS>VHAKHCNDARESRASRELCPGRWRGWPWSOCVTCNGTQERDPVC-----	1002
QY	715	LLARDQCNL--HRKQPELDFCVLRPC	738
Db	1003	RTADDSFGICQERPETARTCLRGPC	1028

RESULT 15

AT53_HUMAN	
AC	AT53_HUMAN STANDARD; PRT; 1205 AA.
AD	015072; 09BX28;
DT	16-OCT-2001 (Rel. 40, Created)
DT	15-JUN-2002 (Rel. 41, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 3) (ADAM-TS3) (Procollagen II
DE	amino-propeptide processing enzyme) (Procollagen II N-proteinase) (PC
DE	II-NP).
GN	ADAMTS3 OR KIAA0366.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RP	[1]
RN	SEQUENCE OF 1-227 FROM N.A.
RX	MEDLINE=21402912; PubMed=11408482;
RX	Fernandes R.J., Hirohata S., Engle J.M., Colige A., Cohn D.H.,
RA	Eyre D.R., Apte S.S.;
RT	"Procollagen II amino propeptide processing by ADAMTS-3. Insights on
RT	dermatosparaxis".
RL	J. Biol. Chem. 276:31502-31509 (2001).
[2]	
RP	SEQUENCE OF 5-1205 FROM N.A.
RP	TISSUE=Brain;
RC	MEDLINE=97349984; PubMed=9205841;
RX	Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RX	Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RA	"Prediction of the coding sequences of unidentified human genes. VII.
RT	The complete sequences of 100 new cDNA clones from brain which can
RT	code for large proteins in vitro.";
RL	DNA Res. 4:141-150(1997).
-!	FUNCTION: Cleaves the propeptides of type II collagen prior to
CC	fibril assembly. Does not act on types I and III collagens.
CC	-! COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC	-! SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC	matrix (by similarity).
CC	-! TISSUE SPECIFICITY: Found in cartilage and skin.
CC	-! DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC	FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC	-! PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC	SIMILARITY).
CC	-! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC	-! SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC	-! SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC	-! CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC	
ENBL	AF247658; AAK28400.1; -
DRDR	ENBL; AB002364; BAA20821.1; -
DRDR	MEROPS; M12.220; -
DRDR	Genew; HGNC:1219; ADAMTS3.
DRDR	MIM; 605011; -
DRDR	InterPro; IPR001762; Disintegrin

DR	InterPro: IPR002870; Pep_M12B_propep.	
DR	InterPro: IPR001590; Reprolysin.	
DR	InterPro: IPR000884; TSP1.	
DR	InterPro: IPR000130; Zn_MTpeptdse.	
DR	Pfam: PF00090; tsp-1; 4.	
DR	Pfam: PF01421; Reprolysin; 1.	
DR	Pfam: PF01562; Pep_M12B_propep; 1.	
DR	SMART: SM00209; TSP1; 4.	
DR	PROSITE: PS50215; ADAM_MEPRO; 1.	
DR	PROSITE: PS50092; TSP1; 2.	
DR	PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.	
DR	PROSITE: PS00427; DISINTEGRIN-1; FALSE NEG.	
DR	Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;	
KW	Repeat; Extracellular matrix; Heparin-binding.	
FT	SIGNAL	1 20
FT	PROPEP	21 249
FT	CHAIN	250 1205
FT	METAL	398 398
FT	ACT_SITE	399 399
FT	METAL	402 402
FT	METAL	408 408
FT	DOMAIN	470 550
FT	DOMAIN	551 607
FT	DOMAIN	608 712
FT	DOMAIN	713 844
FT	DOMAIN	845 902
FT	DOMAIN	903 965
FT	DOMAIN	966 1017
FT	DOMAIN	246 249
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT	CARBOHYD	814 814
FT	CARBOHYD	942 942
FT	CARBOHYD	83 83
FT	CARBOHYD	119 119
FT	CARBOHYD	242 242
FT	CARBOHYD	345 345
FT	CARBOHYD	475 475
FT		

Query Match

Query Match	26.3%;	Score 1062.5;	DB 1;	Length 1205;
Best Local Similarity	33.0%;	Pred. No. 1.4e-69;		
Matches 273;	Conservative 110;	Mismatches 279;	Indels 165;	Gaps 36;
QY	8	VETLVVADESVMYKFGAD-LEHYLLTLTATAARLYRHPHSILNPINIVVVKVLLLRDRDSG	66	
Db	258	IEVLGVDDSDVVRVFGKEHVQNYLLLTLMNIVEIYHDESGLGHINVLVRLMGLCYAKSI	317	
QY	67	PKV-TGNAALTLRNFCANQKLNKLYSDKHPEYWDTAILLFTRODICLGATTCTFLGMADVGT	125	
Db	318	SILIERGNSRSLNVCWRWASQOORSDLNHSHEHDAIFLTLRODFGPA---GMQGYAPVTG	374	
QY	126	MCDPKRSCVIEDDPLSAFTTAHELGHVFNPHD---NVRVCEEVFGKLRANHHMSPTL	182	
Db	375	MCHPYRSCTLNHEDGFSFAFVVAHETGHVLGMEHDGQGNRCGETAMGS-----VMAPLV	429	
QY	183	IQIDRANPWSACSAAILTFDLSHGDCLLDQP---SKPISLPEDLPQASVYLSQOCELA	239	
Db	430	QAAPHRYHWSKSGOELKRYLHS---YDCLLDDPFDHWP-LKPE-LPGINYSMDEQCRFD	485	
QY	240	FGVSKPCPYMQY---CTKLWCTKAKQMVCQTRHFPWADGTSGCGEKLCKKACGVHRH	296	
Db	486	FGVGKMCATFRTDPCKQLWCS-HPDNPFYCKTKGPPLDGTCEAAGKWCYKHCXWK-	543	
QY	297	NLNKRVDSGWAKWDYPGCSRTGGGVQLARRCTNPTTPANGKYGCEGVRVKYSRNL	356	
Db	544	NANQOKDGNWGSWTKFGCSRTCGTYRFRTRCQNPMPINGGQDQCPGVNFYQLCNT	603	
QY	357	PCPSASGKSFREQEAFNG---YNHSTNLTILAVAWPKYSGVSPRDXCKKLCIRANGT	413	
Db	604	ECQKHE--EDFRAQQCQORNHFYQNTKHH-----WLP-YERPDPKKCHLYCQSKET	654	
QY	414	GYFVVLAPKVVDGTLCSS-PDSTSVCVQKCIKAGCDGNLGSKRFRDRCGCVGGDNKSKK	472	
Db	655	GDVAYMKQLVHDGTHCSYKDPYSICIVRGCEYKVGCDKEIGSNKVEDKCGCVGGDNSHCRT	714	

Fri May 2 10:41:54 2003

```
QY 473 VTGLTKPMH--GYNFWAIPAGASSIDIRQGYKGLIGDD----NYLAKN-SQCKYLL 525
Db 715 VKGTFTTRPKUGLYLKMFDIPPGARHV-----LIQEDASPHILAIAKQATGHYIL 765
QY 526 NGH-----FVVSVERDLVVKSGLLRYSGTGTAVESLQASRPLEPLTVEVLVSQKM 577
Db 766 NGKGEAKSRTFDLGVENDYNIEDD-----IESLHTDGPLHDPVIVLIIPQEND 815
QY 578 TPRVRYSFYLPKEPRDKSSHKDPGRGPSVLHNSV-----LSLSNQVEOP--- 623
Db 816 TRSSLTYKYII---HEDSV-----PTINSNNVIQEBLDTFEWALKSWQSVKPCGG 863
QY 624 -----DDRP-PAR-----WVAGSWGPCASCG- 644
Db 864 GFQYTKYGCRRKSDNKMVHRSCFCEANKKXPPIRMCNIECTHPLWVAEEWEHCTKTCGS 923
QY 645 SGLQKRAVDCRGAGQRTVPACDAHRPVETQAC-GE-----PCPT-WELSAWS 691
Db 924 SGYQLRTVRC-----LQPLDGTNRSHVSKYCMGDRPESRRPCNRVPCPAQWKTPWS 976
QY 692 PCSKSCGRGFORSLKCVHGGRLARDQCNLHRKPQELDFCVLRPC 738
Db 977 ECSVTCGEGTEVRQVLC-----RAGDHCD-GEKPESVRACQLPPC 1015
```

Search completed: April 29, 2003, 17:14:25
Job time : 26.0892 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 19:21:51 ; Search time 7406 Seconds
(without alignments)
11211.221 Million cell updates/sec

Title: US-10-009-332-2

Perfect score: 2853

Sequence: 1 atgtctttgtggcatcct.....gcgtctgagccgtgctga 2853

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2848.2	99.8	2853	9	HSA315733	AJ315733 Homo sapi
2	2814.4	98.6	2937	6	AX319860	AX319860 Sequence
3	2812.8	98.6	2930	6	AX342635	AX342635 Sequence
4	957.2	33.6	157963	9	AP002986	AP002986 Homo sapi
5	957.2	33.6	172905	2	AC025130	AC025130 Homo sapi
6	893	31.3	170682	2	AC023429	AC023429 Homo sapi
7	779.8	27.3	182656	2	AC101990	AC101990 Mus muscu
8	778.8	27.3	28000	9	AP003459	AP003459 Homo sapi
9	775.8	27.2	178764	2	AC126507	AC126507 Rattus no
10	775.2	27.2	3927	10	BC009667	BC009667 Mus muscu
11	666	23.3	2670	6	E5282	E52822 Novel metal
12	666	23.3	2670	6	E58655	E58655 Novel metal
13	666	23.3	3711	9	AF060153	AF060153 Homo sapi
14	636	22.3	4180	10	D67076	D67076 Mouse mRNa
15	631.4	22.1	4659	9	AF207664	AF207664 Homo sapi
16	629.8	22.1	3430	9	AF060152	AF060152 Homo sapi
17	629.8	22.1	4014	6	E29406	E29406 Novel integ
18	629.8	22.1	4447	9	AF170084	AF170084 Homo sapi
19	628.8	22.0	2184	6	E29966	E29966 Human ADAMT
20	628.2	22.0	2853	6	E58656	E58656 Novel metal
21	628.2	22.0	4670	9	BC036515	BC036515 Homo sapi
22	626.6	22.0	4309	9	AB037767	AB037767 Homo sapi
23	614.4	21.5	4878	10	AF149118	AF149118 Rattus no
24	608	21.3	2857	10	AF304446	AF304446 Rattus no
25	599	21.0	178764	2	AC126507	AC126507 Rattus no
26	593.6	20.8	3638	10	AF175282	AF175282 Mus muscu
27	584.8	20.5	4193	9	AF148213	AF148213 Homo sapi
28	584.8	20.5	4301	9	AB014588	AB014588 Homo sapi
29	583.2	20.4	4407	6	AX376284	AX376284 Sequence
30	564.2	19.8	3321	10	BC027773	BC027773 Mus muscu
31	547.4	19.2	3002	10	AF140673	AF140673 Mus muscu
32	547.4	19.2	5533	9	AF142099	AF142099 Homo sapi
33	419.2	14.7	5139	9	AB037733	AB037733 Homo sapi
34	419.2	14.7	5808	6	AX319857	AX319857 Sequence
35	417.6	14.6	3674	9	AF261918	AF261918 Homo sapi
36	404	14.2	1251	10	AB042271	AB042271 Rattus no
37	384	13.5	2838	9	HSB802617	AL162080 Homo sapi
38	377.2	13.2	5724	6	AX392458	AX392458 Sequence
39	377.2	13.2	6060	6	AX392460	AX392460 Sequence
40	374	13.1	172905	2	AC025130	AC025130 Homo sapi
41	366.2	12.8	1560	9	AF141293	AF141293 Homo sapi
42	322	11.3	2554	9	AK094057	AK094057 Homo sapi
43	321.2	11.3	1365	10	AB042272	AB042272 Rattus no
44	314.4	11.0	3308	6	AX480930	AX480930 Sequence
45	314.4	11.0	4518	6	AX319858	AX319858 Sequence

ALIGNMENTS

RESULT 1	HSA315733	2853 bp	mRNA	linea	ARI 01-MAR-2002
LOCUS	Homo sapiens	mRNA for metalloprotease disintegrin 15 (ADAMTS15 gene)			
DEFINITION	aj315733				
ACCESSION	aj315733.1	GI:19171175			
VERSION	ADAMTS15	gene; disintegrin; metalloprotease; thrombospondin.			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1				
AUTHORS	Cal.S., Obaya,A.J., Llamazares,M., Garabaya,C., Quesada,V. and Lopez-Otin,C.				

TITLE Cloning, expression analysis, and structural characterization of seven novel human ADAMTS, a family of metalloproteinases with disintegrin and thrombospondin-1 domains

JOURNAL Gene 283 (1-2), 49-62 (2002)

MEDLINE 21856482

PUBMED 11867212

REFERENCE 2 (bases 1 to 2853)

AUTHORS Cal.S.

TITLE Direct Submission

JOURNAL Submitted (26-JUN-2001) Cal S., Biochemistry and Molecular Biology, University of Oviedo, Campus del Cristo, Asturias. 33006, SPAIN

FEATURES

source

1..2853

/organism="Homo sapiens"

/db_xref="taxon:9606"

1..2853

/gene="ADAMTS15"

1..2853

/gene="ADAMTS15"

/codon_start=1

/product="metalloprotease disintegrin 15 with thrombospondin domains"

/protein_id="CAC86014.1"

/db_xref="GI:19171176"

/translation="MELIGILTLAPAGTAGSEPEREVVPIRLDPDINGRYWRG
PEDSGQGLIFQITAFQEDFYHLPLPDAQFLAPARSTELHGLVPLQGLGGSDLRRCF
YSGDVNREDSFAEYSLCGLRGAFGYGAEXVISPFPNASAPAAORNSQGAHLQRR
GVPGVPSADPTSCRGVAGNPAIIRALDPYPRRAGGEGSRSSRRGRKAFVPIPR
YVETLVADDSVMKFGADLEHLLTLAARLYRHPSILNPINIVVVLRLDRD
SGPKVTGAALTLRNCFAWOKLNKVDKPEYWDTAILETRQDLCGACTCDTLGMAD
VTGMDPKRSCSVIEDDGLPSAFTAHGLGHVFNPHONKVCSEVFCKLRANHMSP
TLIQRANPWSACNAIITDFDSHGDCDLDQSKFLSLPELPLGASVTLISQCEL
ATGVGSKPCTQWYCTKLWCTGAKGQWVQTRHPFPWADGTSCEGKICLKGACVERH
NLNKHVDSWAKWDYGPCSRTCGGVOLARRQNTNPANGRYKRGVRYKRSN
LEFPPSSAGSKFRECEAFNGYHNSTNRLTAVAWPKYSGVSPRCKCLICRANG
TGYFYLAEPVVDGLCSPDSTSVQVQKCIKAGDGNLGSKKREDKGVGGDNKSC
KVTGLTPKPMHYNFVAIPAGASSIDIRORGYKLLGDONYLALKNSQKYLNLGH
FVVSVERDLVVKGLLYSGTGTAVESLQASRPILRPLVTLVSVGKMTPPRVRYSF
YLPKEPREDKSHHPDRPGPSVLHNSVLSNQVSPDDPPPARVWAGVSGWPCASGSC
SLQKRVHGRSAGORTVPACDAHRPVETQACGPEPTWELSAWSPCSKSGRGFO
RNSLKVGVHGRLLARDQCNLRKFDCLDFVLRPC"

BASE COUNT 522 a 921 c 903 g 507 t

ORIGIN

Query Match 99.8%; Score 2848.2; DB 9; Length 2853

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2850; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCTTTGCTGGGCATCTTAACCTGGCTTTCCGCGGGCGAACCCGCTGGAGGCTTTGAG 60

DB 1 ATGCTTCTGCTGGGCATCTTAACCTGGCTTTCCGCGGGCGAACCCGCTGGAGGCTTTGAG 60

QY 61 CCAGAGGGGAGTAGTGTTCCTCCAGTCCGAGTGGACCCCGGACATTAACGGCCGCCGTAC 120

DB 61 CCAGAGGGGAGTAGTGTTCCTCCAGTCCGAGTGGACCCCGGACATTAACGGCCGCCGTAC 120

QY 121 TACTGGGGGGTCCCGAGGACTCCGGGACTCAGGAGTCAATTTTTCAGATCAGCATTT 180

DB 121 TACTGGGGGGTCCCGAGGACTCCGGGACTCAGGAGTCAATTTTTCAGATCAGCATTT 180

QY 181 CAGGAGGACTTTTACCTACACCTGACGCGGGATGCTCAGTTCTTGGCTCCCGCTTCTCC 240

DB 181 CAGGAGGACTTTTACCTACACCTGACGCGGGATGCTCAGTTCTTGGCTCCCGCTTCTCC 240

QY 241 ACTGAGCATCTGGGCTGCCCTCCAGGGGCTCACGGGGGCTCTTCAGACCTGCGAGCG 300

DB 241 ACTGAGCATCTGGGCTGCCCTCCAGGGGCTCACGGGGGCTCTTCAGACCTGCGAGCG 300

QY 301 TGCTTCTATTCTGGGACGTGAACCCGAGCCGAGCTCGTTTCGCTGCTGTGAGCCTGTGC 360

DB 301 TGCTTCTATTCTGGGACGTGAACCCGAGCCGAGCTCGTTTCGCTGCTGTGAGCCTGTGC 360

QY 361 GGGGGGCTCCGGGAGGCTTTGGCTACCGAGGGCGCGAGTAGTGTATAGCCCGCTGCC 420

DB 361 GGGGGGCTCCGGGAGGCTTTGGCTACCGAGCGCGAGTAGTGTATAGCCCGCTGCC 420

QY 421 AATGCTAGCGCGCGCGCGCAGCGACACACCGAGGGCGACACCTTCTCCAGCGCGG 480

DB 421 AATGCTAGCGCGCGCGCGCAGCGACACACCGAGGGCGACACCTTCTCCAGCGCGG 480

QY 481 GGTGTTCCGGGGGGGCTTTCCGGAGACCCACCTCTCGCTCGGGGGTGGCTTCGGGCTGG 540

DB 481 GGTGTTCCGGGGGGGCTTTCCGGAGACCCACCTCTCGCTCGGGGGTGGCTTCGGGCTGG 540

QY 541 AACCCCGCATCTTACGGGCGCTTGACACCTTTACAAGCCGCGGGCGGGGCTTCGGGGAG 600

DB 541 AACCCCGCATCTTACGGGCGCTTGACACCTTTACAAGCCGCGGGCGGGGCTTCGGGGAG 600

QY 601 AGTGTAGCGCGCGCAGGTCTGGCGCGCAAGCGCTTTCTGTCTATCCCGGCTACGCTG 660

DB 601 AGTGTAGCGCGCGCAGGTCTGGCGCGCAAGCGCTTTCTGTCTATCCCGGCTACGCTG 660

QY 661 GAGACGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 720

DB 661 GAGACGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 720

QY 721 TATCTGCTGACGCTGCTGGCAACGCGCGGCGACTTACCGCCATCCCGAGCATC.TCAAC 780

DB 721 TATCTGCTGACGCTGCTGGCAACGCGCGGCGACTTACCGCCATCCCGAGCATC.TCAAC 780

QY 781 CCCATCAACATCGTGTGGTCAAGGTGCTGCTTCTAGAGATCGTGACTCGGGGCCAAG 840

DB 781 CCCATCAACATCGTGTGGTCAAGGTGCTGCTTCTAGAGATCGTGACTCGGGGCCAAG 840

QY 841 GTCAACGGCAATGCGGCGCTGACGCTGGCAACCTTCTGCTGGCAGAGAAGCTGAAC 900

DB 841 GTCAACGGCAATGCGGCGCTGACGCTGGCAACCTTCTGCTGGCAGAGAAGCTGAAC 900

QY 901 AAGGTGAGTGAACAGACACCCGAGTACTGGGACACTGCCATCCTCTTACACAGCAGGAC 960

DB 901 AAGGTGAGTGAACAGACACCCGAGTACTGGGACACTGCCATCCTCTTACACAGCAGGAC 960

QY 961 CTGTGTGGAGGACACCATGTGTGACACCCCTGGGCGATGGTGTGGTGGTGGTGGTGGTGG 1020

DB 961 CTGTGTGGAGGACACCATGTGTGACACCCCTGGGCGATGGTGTGGTGGTGGTGGTGGTGG 1020

QY 1021 CCCAAGAGAAGTGTGTCTATTGAGGACATGGGCTTCCATCAGGCTTCCACACTGCC 1080

DB 1021 CCCAAGAGAAGTGTGTCTATTGAGGACATGGGCTTCCATCAGGCTTCCACACTGCC 1080

QY 1081 CACGAGCTGGGCGCAGCTGTTCAACATGCCCATGCCATGACATGTGAAAGTCTGTGAGAGGTG 1140

DB 1081 CACGAGCTGGGCGCAGCTGTTCAACATGCCCATGCCATGACATGTGAAAGTCTGTGAGAGGTG 1140

QY 1141 TTTGSGAAGCTCCGAGCAACACATGATGTCCCGGACCCCTCATCCAGATCGACCGTGGC 1200

DB 1141 TTTGSGAAGCTCCGAGCAACACATGATGTCCCGGACCCCTCATCCAGATCGACCGTGGC 1200

QY 1201 AACCCCTGGTTCAGCTTCAGTGTCCCATATCATCACCTTCTCTGGACAGGGGACAGGT 1260

DB 1201 AACCCCTGGTTCAGCTTCAGTGTCCCATATCATCACCTTCTCTGGACAGGGGACAGGT 1260

QY 1261 GACTGCTCTTCGACCAACCCAGCAAGCCCATCTCCCTGCCGCGAGGATCTGCCGGGCGCC 1320

DB 1261 GACTGCTCTTCGACCAACCCAGCAAGCCCATCTCCCTGCCGCGAGGATCTGCCGGGCGCC 1320

QY 1321 AGCTACACCTTGAGCGACAGTGGCTGCTTTTGGCGTGGGCTCCAAAGCCCTGTGCTCT 1380

DB 1321 AGCTACACCTTGAGCGACAGTGGCTGCTTTTGGCGTGGGCTCCAAAGCCCTGTGCTCT 1380

QY 1381 TACATGAGTACTGACCAAGCTGTGGTGCACCGGAGGCCAAGGACAGATGTGTGTGC 1440

DB 1381 TACATGAGTACTGACCAAGCTGTGGTGCACCGGAGGCCAAGGACAGATGTGTGTGC 1440

QY 1441 CAGACCGGCGCACTTCCCTCTGGGCGGATGGGACAGCTGTGGCGAGGCAAGCTGTGCGCTC 1500

Db 1441 CAGACCGCCACTTCCCTCGGGCCGATGGCACCAGCTGTGGCGAGGCAAGCTCTGCCTC 1500
QY 1501 AAAGGGCCCTCGTGGAGACACAACTCAACACACAGGCTGATGGTTCCTGGGCC 1560
Db 1501 AAAGGGCCCTCGTGGAGACACAACTCAACACACAGGCTGATGGTTCCTGGGCC 1560
QY 1561 AAATGGGATCCCTATGAGCCCTCTGCTGGCGACATGTGTGGGGGCGTGCAGCTGGCCAGG 1620
Db 1561 AAATGGGATCCCTATGAGCCCTCTGCTGGCGACATGTGTGGGGGCGTGCAGCTGGCCAGG 1620
QY 1621 AGGAGTGCACCAACCCCTCTGCAAGGGGGGAAGTACTGCGAGGAGTGGGGTG 1680
Db 1621 AGGAGTGCACCAACCCCTCTGCAAGGGGGGAAGTACTGCGAGGAGTGGGGTG 1680
QY 1681 AAATACCGATCTTGCACCTTGGAGCCCTGCCAGCTCAAGCTCCCGAAGAGCTTCCGG 1740
Db 1681 AAATACCGATCTTGCACCTTGGAGCCCTGCCAGCTCAAGCTCCCGAAGAGCTTCCGG 1740
QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCACCAACCGGCTCACTCTCGCC 1800
Db 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCACCAACCGGCTCACTCTCGCC 1800
QY 1801 GTGGCATGGTGGCCCAAGTACTCGGGGTGTCTCCCGGGGACAAGTGCAGCTCATCTGC 1860
Db 1801 GTGGCATGGTGGCCCAAGTACTCGGGGTGTCTCCCGGGGACAAGTGCAGCTCATCTGC 1860
QY 1861 CGAGCCCAATGGCACTGGCTACTTCTATGTGCTGGCACCACCAAGTGTGGACGCGACGCTG 1920
Db 1861 CGAGCCCAATGGCACTGGCTACTTCTATGTGCTGGCACCACCAAGTGTGGACGCGACGCTG 1920
QY 1921 TGCTTCTCTGACTCCAGCTCCGCTGTGTGTCGAAGGCAAGTGCATCAAGGCTGGCTGTGAT 1980
Db 1921 TGCTTCTCTGACTCCAGCTCCGCTGTGTGTCGAAGGCAAGTGCATCAAGGCTGGCTGTGAT 1980
QY 1981 GGGAACTGGCTCCAGAGAGATTCGACAAGTGTGGGTGTGGGGGAGACATAAG 2040
Db 1981 GGGAACTGGCTCCAGAGAGATTCGACAAGTGTGGGTGTGGGGGAGACATAAG 2040
QY 2041 AGCTGCAAGAGGTGACTGGACTCTTCAACCAAGCCATGATGGCTACAAATTTCTGGGTG 2100
Db 2041 AGCTGCAAGAGGTGACTGGACTCTTCAACCAAGCCATGATGGCTACAAATTTCTGGGTG 2100
QY 2101 GCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGAGCGGTTACAAAGGCGTGTATC 2160
Db 2101 GCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGAGCGGTTACAAAGGCGTGTATC 2160
QY 2161 GGGATGACAACTACCTGGCTCTGAAGACAGCAAGCAAGTACCTGCTCAACGGGCAT 2220
Db 2161 GGGATGACAACTACCTGGCTCTGAAGACAGCAAGCAAGTACCTGCTCAACGGGCAT 2220
QY 2221 TTCTGTGTGTGGGGTGGAGCGGACCTGGTGTGAAGGGCAGTCTGCTGGGTACAGC 2280
Db 2221 TTCTGTGTGTGGGGTGGAGCGGACCTGGTGTGAAGGGCAGTCTGCTGGGTACAGC 2280
QY 2281 GGCACGGGCACAGGGTGGAGAGCGCTGACGGCTTCGCGGCCCATCTGAGCGCGCTGACC 2340
Db 2281 GGCACGGGCACAGGGTGGAGAGCGCTGACGGCTTCGCGGCCCATCTGAGCGCGCTGACC 2340
QY 2341 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTATCTG 2400
Db 2341 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTATCTG 2400
QY 2401 CCCAAAGAGCTCGGGAGGACAAAGTCTCTCATCCCAAGGACCCCGGGGACCTCTCTG 2460
Db 2401 CCCAAAGAGCTCGGGAGGACAAAGTCTCTCATCCCAAGGACCCCGGGGACCTCTCTG 2460
QY 2461 TTGCACAAAGCGTCTCAGCCCTCTCCAAAGGAGAGCGGAGCGGACGAGCGCCCT 2520
Db 2461 TTGCACAAAGCGTCTCAGCCCTCTCCAAAGGAGAGCGGAGCGGAGCGGACGAGCGCCCT 2520
QY 2521 GCAGGCTGGTGGCTGGCAGCTGGGGCGGTGCTCCCGAGCTGCGGCGAGTGGCCCTGCAG 2580
Db 2521 GCAGGCTGGTGGCTGGCAGCTGGGGCGGTGCTCCCGAGCTGCGGCGAGTGGCCCTGCAG 2580

QY 2581 AAGCGCGGTGGACTCGCGGGCTCCGCGGGCAGCGATCCCTGCTGTGATGCA 2640
Db 2581 AAGCGCGGTGGACTCGCGGGCTCCGCGGGCAGCGATCCCTGCTGTGATGCA 2640
QY 2641 GCGCATCGCGCGTGGAGACACAAAGCTTGGGGAGCGCTGCCACCTGGGAGCTCAGC 2700
Db 2641 GCGCATCGCGCGTGGAGACACAAAGCTTGGGGAGCGCTGCCACCTGGGAGCTCAGC 2700
QY 2701 GCCTGGTCACTGCTTCAAGAGCTGCGCGGGGAGTTCAGAGCGCTCACTCAAGTGT 2760
Db 2701 GCCTGGTCACTGCTTCAAGAGCTGCGCGGGGAGTTCAGAGCGCTCACTCAAGTGT 2760
QY 2761 GTGGCCCAAGGCGGCTGCTGCGCGGGACCAAGTGCACCTGACCGCAAGCCCGAG 2820
Db 2761 GTGGCCCAAGGCGGCTGCTGCGCGGGACCAAGTGCACCTGACCGCAAGCCCGAG 2820
QY 2821 GAGCTGGACTTCTGCTGCTCTGAGCGCGTGTGA 2853
Db 2821 GAGCTGGACTTCTGCTGCTCTGAGCGCGTGTGA 2853

RESULT 2
AX319860
LOCUS AX319860 2937 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 24 from Patent WO0183782.
ACCESSION AX319860
VERSION AX319860.1 GI:17901450
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Plowman, G.D., Whyte, D., Sudarsanam, S., Manning, G., Cuenepiel, S. and Payne, V.
TITLE Novel proteases
JOURNAL Patent: WO 0183782-A 24 08-NOV-2001;
Sugen, Inc. (US)
FEATURES
source Location/Qualifiers
1..2937
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 531 a 958 c 936 g 512 t
ORIGIN

Query Match 98.6%; Score 2814.4; DB 6; Length 2937;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2847; Conservative 0; Mismatches 6; Indels 9; Gaps 2;

QY 1 ATGCTTTTGTGGGCATCCTAACCTGGCTTTCGCCGGCGAACCGCTGGAGGCTTTCAG 60
Db 76 ATGCTTCTGTGGGCATCCTAACCTGGCTTTCGCCGGCGAACCGCTGGAGGCTTTCAG 135
QY 61 CCAGAGCGGGAGGTAGTCTTCCCATCCGACTGGACCGGACATTAACGGCGCGCTAC 120
Db 136 CCAGAGCGGGAGGTAGTCTTCCCATCCGACTGGACCGGACATTAACGGCGCGCTAC 195
QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCAGACATTT 180
Db 196 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCAGACATTT 255
QY 181 CAGGAGGACTTTTACTACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCTTCC 240
Db 256 CAGGAGGACTTTTACTACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCTTCC 315
QY 241 ACTGAGCATCTGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGAGCTCGGACG 300
Db 316 ACTGAGCATCTGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGAGCTCGGACG 375
QY 301 TCCTTCTATCTGGGGAGTGAACCGGAGCGGACTCGTTCGGCTGCTGTGAGCTGTGC 360
Db 376 TCCTTCTATCTGGGGAGTGAACCGGAGCGGACTCGTTCGGCTGCTGTGAGCTGTGC 435

QY 361 GGGGGGCTCCGGAGAGCCTTTGGCTACCGAGCGCCGAGTATGTCATATTAGCCCGCTGCCCC 420
Db 436 GGGGGGCTCCGGAGAGCCTTTGGCTACCGAGCGCCGAGTATGTCATATTAGCCCGCTGCCCC 495
QY 421 AATGCTAGCGCGCGCGCGCGAGCCACAGACCGAGCGGCACACCTTCTCCAGCGCCGG 480
Db 496 AATGCTAGCGCGCGCGCGCGAGCCACAGACCGAGCGGCACACCTTCTCCAGCGCCGG 555
QY 481 GGTGTTCCGGCGCGCGCTTCCGAGAGCCACACCTCTCGTCTCGGGGTGGCGCTCGG 540
Db 556 GGTGTTCCGGCGCGCGCTTCCGAGAGCCACACCTCTCGTCTCGGGGTGGCGCTCGG 615
QY 541 AACCCCGCCATCTACGGGCCCTTGGACCCCTTACAACCGCGCGCGCGCGCTTCCGGGAG 600
Db 616 AACCCCGCCATCTACGGGCCCTTGGACCCCTTACAACCGCGCGCGCGCGCTTCCGGGAG 675
QY 601 AGTCGTAGCGCGCGAGTCTGGCGCGCGCCCAAGCGTTTCTGTCTATCCCGGGTACGTG 660
Db 676 AGTCGTAGCGCGCGAGTCTGGCGCGCGCCCAAGCGTTTCTGTCTATCCCGGGTACGTG 735
QY 661 GAGACGCTGTGTGTCGGGACGAGTCAATGGTCAAGTCTCCACGCGCGGACCTTGAACAT 720
Db 736 GAGACGCTGTGTGTCGGGACGAGTCAATGGTCAAGTCTCCACGCGCGGACCTTGAACAT 795
QY 721 TATCTGCTGACGCTGTGGCAACGGCGCGGAGTCTACGGCCATCCAGATCTCTCAAC 780
Db 796 TATCTGCTGACGCTGTGGCAACGGCGCGGAGTCTACGGCCATCCAGATCTCTCAAC 855
QY 781 CCATCAACATCTGTGGTCAAGTGTCTCTTCTAGAGATCGTCACTCCGGGCCCAAG 840
Db 856 CCATCAACATCTGTGGTCAAGTGTCTCTTCTAGAGATCGTCACTCCGGGCCCAAG 915
QY 841 GTCACCGGCAATCGGCCCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAGCTGAAC 900
Db 916 GTCACCGCAATCGGCCCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAGCTGAAC 975
QY 901 AAAGTGAGTACAAAGACCCCGAGTACTGGGACACTGCCATCTCTTACCAGGCGAGGAC 960
Db 976 AAAGTGAGTACAAAGACCCCGAGTACTGGGACACTGCCATCTCTTACCAGGCGAGGAC 1035
QY 961 CTGTGTGGAGCCACCACTGTGACACCTTGGCATGGCTGATGTGGGTACCATGTGTGAC 1020
Db 1036 CTGTGTGGAGCCACCACTGTGACACCTTGGCATGGCTGATGTGGGTACCATGTGTGAC 1095
QY 1021 CCCAAGAGAGTGTCTGTCTATTGAGGAGATGGGCTTCCATCAGCCTTCAACATGCC 1080
Db 1096 CCCAAGAGAGTGTCTGTCTATTGAGGAGATGGGCTTCCATCAGCCTTCAACATGCC 1155
QY 1081 CAGAGCTGGGCGACGTTTCAACATGCCCATGCCATGAAATGTGAGGAGGTG 1140
Db 1156 CAGAGCTGGGCGACGTTTCAACATGCCCATGCCATGAAATGTGAGGAGGTG 1215
QY 1141 TTTGGGAAGCTCCGAGCAACCACTGATGTCCCGACCTTATCCAGATCCAGCCTGCC 1200
Db 1216 TTTGGGAAGCTCCGAGCAACCACTGATGTCCCGACCTTATCCAGATCCAGCCTGCC 1275
QY 1201 AACCCCTGTGACGTGATGTGATCATCATCAGACTTCTGTGACAGCGGCGACGCT 1260
Db 1276 AACCCCTGTGACGTGATGTGATCATCATCAGACTTCTGTGACAGCGGCGACGCT 1335
QY 1261 GACTGCTCTTGACCAACCCAGACCCATCTCCCTGCCCGAGGATCTGCCGGCGCC 1320
Db 1336 GACTGCTCTTGACCAACCCAGACCCATCTCCCTGCCCGAGGATCTGCCGGCGCC 1395
QY 1321 AGCTACACCTGAGCAGCAGTCCGAGCTGGCTTTTGGCGTGGGCTTCCAAAGCCCTGTCT 1380
Db 1396 AGCTACACCTGAGCAGCAGTCCGAGCTGGCTTTTGGCGTGGGCTTCCAAAGCCCTGTCT 1455
QY 1381 TACATGAGTGTGACCAAGCTGTGTGACCGGGAAGGCCAAGGACAGATGTGTGC 1440
Db 1456 TACATGAGTGTGACCAAGCTGTGTGACCGGGAAGGCCAAGGACAGATGTGTGC 1515

QY 1441 CAGACCCCGCACTTCCCTGGCCGATGGCACAGCTGTGGCGAGGGCAAGCTCTGCCTC 1500
Db 1516 CAGACCCCGCACTTCCCTGGCCGATGGCACAGCTGTGGCGAGGGCAAGCTCTGCCTC 1575
QY 1501 AAAGGGGCTCGTGGAGAGACAAACCTCAACAGACAGGGTGGATGTTCTCTGGGCC 1560
Db 1576 AAAGGGGCTCGTGGAGAGACAAACCTCAACAGACAGGGTGGATGTTCTCTGGGCC 1635
QY 1561 AAATGGGATCCTATGGCCCTTGTCTCGGCACATGTGTGGGGGGCTGCAGCTGGCCAGG 1620
Db 1636 AAATGGGATCCTATGGCCCTTGTCTCGGCACATGTGTGGGGGGCTGCAGCTGGCCAGG 1695
QY 1621 AGGAGTGCACCAACCCCAACCCCTGCAACGGGGGCAAGTACTCGGAGGAGTGAAGGTG 1680
Db 1696 AGGAGTGCACCAACCCCAACCCCTGCAACGGGGGCAAGTACTCGGAGGAGTGAAGGTG 1755
QY 1681 AAATACCATCTCTCAACCTGGAGCCCTGCGCCAGCTCAGCTCCGGAAGAGCTTCGG 1740
Db 1756 AAATACCATCTCTCAACCTGGAGCCCTGCGCCAGCTCAGCTCCGGAAGAGCTTCGG 1815
QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTTACAACACAGCACACCCGCTCACTCTCGCC 1800
Db 1816 GAGGAGCAGTGTGAGGCTTTCAACGGCTTACAACACAGCACACCCGCTCACTCTCGCC 1875
QY 1801 GTGGCATGGTGGCCCAAGTACTCCGGCGTGTCTCCCGGGACAAAGTCAAGCTCATCTGC 1860
Db 1876 GTGGCATGGTGGCCCAAGTACTCCGGCGTGTCTCCCGGGACAAAGTCAAGCTCATCTGC 1935
QY 1861 CGAGCAATGGCAGTGTCTTCTATGTCTGGCACCCAA---GGTGGTGGAGCGGACG 1917
Db 1936 CGAGCAATGGCAGTGTCTTCTATGTCTGGCACCCAAAGTGTGGTGGAGCGGACG 1995
QY 1918 CTGTGCTCTCTGACTCCACCTCCGCTGTGTCCAAGCAAGTGCATCAAGGCTGGGTGT 1977
Db 1996 CTGTGCTCTCTGACTCCACCTCCGCTGTGTCCAAGCAAGTGCATCAAGGCTGGGTGT 2055
QY 1978 GATGGAACTGGGCTCCAAAGAGATTCCAGAGTGTGGGTGTGGGGGAGACAAT 2037
Db 2056 GATGGAACTGGGCTCCAAAGAGATTCCAGAGTGTGGGTGTGGGGGAGACAAT 2115
QY 2038 AAGAGCTGCAAGAGTGTGACTCTTCAACAGCCCATGCATGGCTACAATTTCTGTG 2097
Db 2116 AAGAGCTGCAAGAGTGTGACTCTTCAACAGCCCATGCATGGCTACAATTTCTGTG 2175
QY 2098 GTGGCCATCCCGCAGGCGCTCAAGATCGACATCCGCAAGCGGTTTCAAGAGGCTG 2157
Db 2176 GTGGCCATCCCGCAGGCGCTCAAGATCGACATCCGCAAGCGGTTTCAAGAGGCTG 2235
QY 2158 ATCGGGGATGACAATCTGCTCTGAAGAACAGCAAGCAAGTACCTCTCAACGGG 2217
Db 2236 ATCGGGGATGACAATCTGCTCTGAAGAACAGCAAGCAAGTACCTCTCAACGGG 2295
QY 2218 CATTTCTGTTGCTCGGCGGTGAGCGGACCTGTGTGAAGGCGAGTCTGCTCGGTAC 2277
Db 2296 CATTTCTGTTGCTCGGCGGTGAGCGGACCTGTGTGAAGGCGAGTCTCTCGGTAC 2355
QY 2278 AGCGCACGGGCACAGCGGTGAGAGCCTGAGGCTTCCCGGCCCATCTCTGGAGCGGCTG 2337
Db 2356 AGCGCACGGGCACAGCGGTGAGAGCCTGAGGCTTCCCGGCCCATCTCTGGAGCGGCTG 2415
QY 2338 ACCGTGAGGTCTCTCCGTGGGAAGATGACCCGCCGGGTCCGCTACTCTCTTAT 2397
Db 2416 ACCGTGAGGTCTCTCCGTGGGAAGATGACCCGCCGGGTCCGCTACTCTCTTAT 2475
QY 2398 CTGCCAAAGAGCTTCGGGAGGACAAGTCTCTCATCCCAAGGACCCCGCG- ----GGA 2451
Db 2476 CTGCCAAAGAGCTTCGGGAGGACAAGTCTCTCATCCCGCGCACCCCGGGAGGAGGA 2535
QY 2452 CCCTCTGTGTCACAAACAGCGTCTCAGCCTCTTCAACACAGGTGGAGCGCGGACGAC 2511
Db 2536 CCCTCTGTGTCACAAACAGCGTCTCAGCCTCTTCAACACAGGTGGAGCGCGGACGAC 2595
QY 2512 AGGCCCCCTGCACGCTGGGTGGCTGGCAGCTGGGGCGCGTGTCTCCGCGAGCTGGCGCAGT 2571

```
Db 2596 AGGCCCCCTGACAGCTGGGTGGTGGAGCTGGGGGCGCTGCTCCGCGAGCTGCGGAGT 2655
Qy 2572 GGCCTGCAGAAAGGGGGGTGGACTGCGGGGCTCCGCGGCGAGCGCAGGTCCTGCGC 2631
Db 2656 GGCCTGCAGAAAGGGGGGTGGACTGCGGGGCTCCGCGGCGAGCGCAGGTCCTGCGC 2715
Qy 2632 TGTGATCAGCCATCCGCGGCGGTGGAGACACAAAGCCTGCGGGGAGCCCTGCCACACTGG 2691
Db 2716 TGTGATCAGCCATCCGCGGCGGTGGAGACACAAAGCCTGCGGGGAGCCCTGCCACACTGG 2775
Qy 2692 GAGCTAGCGGCTGGTACCTGCTCCAGAGCTGCGGGGCGGGGATTTGAGAGGCGCTCA 2751
Db 2776 GAGCTAGCGGCTGGTACCTGCTCCAGAGCTGCGGGGCGGGGATTTGAGAGGCGCTCA 2835
Qy 2752 CTCAAGTGTGTGGCCACGAGGCGGCTGCTGCGCCGCGGACAGTGCACCTTGCACCG 2811
Db 2836 CTCAAGTGTGTGGCCACGAGGCGGCTGCTGCGCCGCGGACAGTGCACCTTGCACCG 2895
Qy 2812 AAGCCCGAGAGTGGACTTCTCGCTCCTGAGCCGCTGTGA 2853
Db 2896 AAGCCCGAGAGTGGACTTCTCGCTCCTGAGGCGCTGTGA 2937

RESULT 3
AX342635
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
AUTHORS
Yue H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, J.,
Tribouley, C.M., Deleage, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A.,
Hafalia, A., Khan, F.A., Wallia, N.K., Yao, M.G., Lu, D.A., Patterson, C.,
Tang, Y.T., Walsh, R.T., Azimzai, Y., Ramkumar, J., Xu, Y. and Reddy, R.
JOURNAL
Patent: WO 0198468-A 32 27-DEC-2001;
Incyte Genomics, Inc. (US)
FEATURES
Source
1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 7473089CB1"
BASE COUNT
529 a 958 c 931 g 512 t
ORIGIN
Query Match 98.6%; Score 2812.8; DB 6; Length 2930;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2843; Conservative 0; Mismatches 7; Indels 6; Gaps 2;
Qy 1 ATGCTTTTCTGGGATCTTAACCTTGGCTTTCGCGGCGGAGACGCTGGAGCTTTGAG 60
Db 75 ATGCTTCTGCTGGGATCTTAACCTTGGCTTTCGCGGCGGAGACGCTGGAGCTTTGAG 134
Qy 61 CCAGAGCGGAGGTAGTCTTCCATCCGACTGGACCGCGGACATTAAGCGCGCGGCTAC 120
Db 135 CCAGAGCGGAGGTAGTCTTCCATCCGACTGGACCGCGGACATTAAGCGCGCGGCTAC 194
Qy 121 TACTGCGGGTCCCGAGAGCTCCGGGATCAGGACTCATTTTTCAGATCACAGATTT 180
Db 195 TACTGCGGGTCCCGAGAGCTCCGGGATCAGGACTCATTTTTCAGATCACAGATTT 254
Qy 181 CAGGAGGACTTTTACTACACTGACGCGGATGCTCAGTTCTTGGCTCCGCGCTTCC 240
Db 255 CAGGAGGACTTTTACTACACTGACGCGGATGCTCAGTTCTTGGCTCCGCGCTTCC 314
Qy 241 ACTGAGACTTCTGGGCTCCCGCTCCAGGGGCTTCCAGGGGCTTCTTTCAGACTGCGAGCG 300
Db 315 ACTGAGACTTCTGGGCTCCCGCTCCAGGGGCTTCCAGGGGCTTCTTTCAGACTGCGAGCG 374
```

```
Qy 301 TGCCTTCTATCTGGGGAGCTGAACCGCGAGCGGAGCTGTTCCGCTGCTGTGAGCGTGTGC 360
Db 375 TGCCTTCTATCTGGGGAGCTGAACCGCGAGCGGAGCTGTTCCGCTGCTGTGAGCGTGTGC 434
Qy 361 GGGGGGCTCCGCGGAGCGCTTTGGCTTACCAGAGGCGCGGAGTATGTCATTAGCCGCTGCC 420
Db 435 GGGGGGCTCCGCGGAGCGCTTTGGCTTACCAGAGGCGCGGAGTATGTCATTAGCCGCTGCC 494
Qy 421 AATGTAGCGCGCGCGGCGGCGAGCGCAACAGCCAGGCGGCACACCTTCTTCCAGCGCGG 480
Db 495 AATGTAGCGCGCGCGGCGGCGAGCGCAACAGCCAGGCGGCACACCTTCTTCCAGCGCGG 554
Qy 481 GGTGTTCCGGGGGCGCTTCCGGAGACCCCACTTCTGCTGCGGGGTGGCCCTCGGGCTGG 540
Db 555 GGTGTTCCGGGGGCGCTTCCGGAGACCCCACTTCTGCTGCGGGGTGGCCCTCGGGCTGG 614
Qy 541 AACCCCGCATCTTACGGGCGCTTGAACCTTACAAGCGCGCGGCGGCGGCTTCCGGGAG 600
Db 615 AACCCCGCATCTTACGGGCGCTTGAACCTTACAAGCGCGCGGCGGCGGCTTCCGGGAG 674
Qy 601 AGTCGTAGCGCGCGGCGGCGAGGTCTGGGCGCGCCAAAGCGTTTCTGTCATCCCGGCTAGCGT 660
Db 675 AGTCGTAGCGCGCGGCGGCGAGGTCTGGGCGCGCCAAAGCGTTTCTGTCATCCCGGCTAGCGT 734
Qy 661 GAGAGCGTGGTGGTCCGCGAGAGTCAATGTGTCAGTTTCCAGCGCGGCGGAGCTTGAACAT 720
Db 735 GAGAGCGTGGTGGTCCGCGAGAGTCAATGTGTCAGTTTCCAGCGCGGCGGAGCTTGAACAT 794
Qy 721 TATCTGTGACGCTGCTGGCAACGCGCGGCGGAGCTTACCAGCATCCAGCATCTCTCAAC 780
Db 795 TATCTGTGACGCTGCTGGCAACGCGCGGCGGAGCTTACCAGCATCCAGCATCTCTCAAC 854
Qy 781 CCCATCAACATCTGTGGTCAAGGTGCTGCTTCTTAGAGATCGTACTCCGGGCGGCAAG 840
Db 855 CCCATCAACATCTGTGGTCAAGGTGCTGCTTCTTAGAGATCGTACTCCGGGCGGCAAG 914
Qy 841 GTCACCGCAATGGGCGCTGACGCTGGGCAACTTCTGTGCTGGCAGAAAGCTCAAC 900
Db 915 GTCACCGCAATGGGCGCTGACGCTGGGCAACTTCTGTGCTGGCAGAAAGCTCAAC 974
Qy 901 AAGTGTAGTGAACAGACCCCGAGTACTGGGACACTGCTCTTCTTCCAGGAGGAG 960
Db 975 AAGTGTAGTGAACAGACCCCGAGTACTGGGACACTGCTCTTCTTCCAGGAGGAG 1034
Qy 961 CTGTGTGGAGCACACCTGTGACACCTGGGCGATGCTGATGTGGTACCATCTGTGAC 1020
Db 1035 CTGTGTGGAGCACACCTGTGACACCTGGGCGATGCTGATGTGGTACCATCTGTGAC 1094
Qy 1021 CCCAAGAGAAGCTGCTGTCTTATTGAGGACGATGGGCTTCCATCAGCTTCCAGGAGTGC 1080
Db 1095 CCCAAGAGAAGCTGCTGTCTTATTGAGGACGATGGGCTTCCATCAGCTTCCAGGAGTGC 1154
Qy 1081 CACGAGCTGGGCGAGTGTTCACATGCCCGCATGACATGAAAGTCTGTGAGGAGGTG 1140
Db 1155 CACGAGCTGGGCGAGTGTTCACATGCCCGCATGACATGAAAGTCTGTGAGGAGGTG 1214
Qy 1141 TTTGGGAAGTCCCGAGCAACACATGATGTCGCCGAGCTTCCATCCAGATCGAGCTGTC 1200
Db 1215 TTTGGGAAGTCCCGAGCAACACATGATGTCGCCGAGCTTCCATCCAGATCGAGCTGTC 1274
Qy 1201 AACCCCTGGTACGCTCGAGTGTGCCATCATCCGACTTCTTGGAGACAGGCGGAGTGC 1260
Db 1275 AACCCCTGGTACGCTCGAGTGTGCCATCATCCGACTTCTTGGAGACAGGCGGAGTGC 1334
Qy 1261 GACTGCTCTTGGAGCAACCGAGCAAGCCATCTTCCCTGCGCGGAGGATGTCGCGGGCGCC 1320
Db 1335 GACTGCTCTTGGAGCAACCGAGCAAGCCATCTTCCCTGCGCGGAGGATGTCGCGGGCGCC 1394
Qy 1321 AGCTACACCTTGGAGCAAGCTGCGAGCTGGCTTTTGGCGTGGGCTCCAGGCGCTGTCT 1380
Db 1395 AGCTACACCTTGGAGCAAGCTGCGAGCTGGCTTTTGGCGTGGGCTCCAGGCGCTGTCT 1454
```

```
QY 1381 TACATCAGTACTGCACCAAGCTGTGGTACCGGGAAGCCAGGACAGATGTTGTC 1440
|||||
Db 1455 TACATCAGTACTGCACCAAGCTGTGGTACCGGGAAGCCAGGACAGATGTTGTC 1514
|||||
QY 1441 CAGACCCGCACTTCCCTGGGCGGANTGCACAGCTGTGGCGAGGGGCAAGCTCTGCCCTC 1500
|||||
Db 1515 CAGACCCGCACTTCCCTGGGCGGANTGCACAGCTGTGGCGAGGGGCAAGCTCTGCCCTC 1574
|||||
QY 1501 ARAAGGGCTGGTGGAGAGACACACCTCAACAAGCAGAGGTTGATGTTCTTGGGCC 1560
|||||
Db 1575 ARAAGGGCTGGTGGAGAGACACACCTCAACAAGCAGAGGTTGATGTTCTTGGGCC 1634
|||||
QY 1561 AAATGGGATCCCTATATGGCCCTCTGTCGCGACATGTGTGGGGGCTGCAGCTTGCCAGG 1620
|||||
Db 1635 AAATGGGATCCCTATATGGCCCTCTGTCGCGACATGTGTGGGGGCTGCAGCTTGCCAGG 1694
|||||
QY 1621 AGCAGTGCACCAACCCACCCCTGCGCAACGGGGGCAAGTACTGCGAGGGAGTGAGGGTG 1680
|||||
Db 1695 AGCAGTGCACCAACCCACCCCTGCGCAACGGGGGCAAGTACTGCGAGGGAGTGAGGGTG 1754
|||||
QY 1681 AAATACCGATCTTGCACCTGGAGCCCTGCCAGCTCAGCCCTCGGAAAGAGCTTCCGG 1740
|||||
Db 1755 AAATACCGATCTTGCACCTGGAGCCCTGCCAGCTCAGCCCTCGGAAAGAGCTTCCGG 1814
|||||
QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGACCAACCGGCTCACTCTCGCC 1800
|||||
Db 1815 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGACCAACCGGCTCACTCTCGCC 1874
|||||
QY 1801 GTGGCATGGTGCCCAAGTACTCCGGCGTGTCTCCCGGGACAAGTGCAAGCTCATCTGC 1860
|||||
Db 1875 GTGGCATGGTGCCCAAGTACTCCGGCGTGTCTCCCGGGACAAGTGCAAGCTCATCTGC 1934
|||||
QY 1861 CGAGCAATGGCACTGGCTACTTCTATGTCTGGCAACCAA---GGTGGTGGACGGCAGC 1917
|||||
Db 1935 CGAGCAATGGCACTGGCTACTTCTATGTCTGGCAACCAAAGGTGTGTGGACGGCAGC 1994
|||||
QY 1918 CTGTGCTCTCTGACTCCAGCTCCGCTGTGTGTCCAAAGGCAAGTGCAATCAAGGCTGCTGT 1977
|||||
Db 1995 CTGTGCTCTCTGACTCCAGCTCCGCTGTGTGTCCAAAGGCAAGTGCAATCAAGGCTGCTGT 2054
|||||
QY 1978 GATGGGAACCTGGCTCCCAAGAGATTCGACAAGTGTGGGTGTGTGGGGAGACAAT 2037
|||||
Db 2055 GATGGGAACCTGGCTCCCAAGAGATTCGACAAGTGTGGGTGTGTGGGGAGACAAT 2114
|||||
QY 2038 AAGAGCTGCAAGAGGTGACTGGACCTTCCACCAAGCCCATGCATGGCTACAATTTCTGTG 2097
|||||
Db 2115 AAGAGCTGCAAGAGGTGACTGGACCTTCCACCAAGCCCATGCATGGCTACAATTTCTGTG 2174
|||||
QY 2098 GTGGCCATCCCGAGCGGCTCAAGATCGACATCGCGAGCGCGGTTACAAGGGCTG 2157
|||||
Db 2175 GTGGCCATCCCGAGCGGCTCAAGATCGACATCGCGAGCGCGGTTACAAGGGCTG 2234
|||||
QY 2158 ATCGGGATCACAACTACCTGGCTCTGAAGACAGCAAGGCAAGTACCTGTCTCAACGGG 2217
|||||
Db 2235 ATCGGGATCACAACTACCTGGCTCTGAAGACAGCAAGGCAAGTACCTGTCTCAACGGG 2294
|||||
QY 2218 CATTTCTGTGTGCGGGTGGAGCGGGACCTGTGTGTGAAGGGCAAGTCTGTCTCGGGTAC 2277
|||||
Db 2295 CATTTCTGTGTGCGGGTGGAGCGGGACCTGTGTGTGAAGGGCAAGTCTGTCTCGGGTAC 2354
|||||
QY 2278 AGCGGCACGGGCACAGGGTGGAGAGCTGCAGGCTTCCGGGCCCATCTCTGGAGCGCGTG 2337
|||||
Db 2355 AGCGGCACGGGCACAGGGTGGAGAGCTGCAGGCTTCCGGGCCCATCTCTGGAGCGCGTG 2414
|||||
QY 2338 ACCGTGAGGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTAT 2397
|||||
Db 2415 ACCGTGAGGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTAT 2474
|||||
QY 2398 CTGCCAAAGAGCTCGGGAGGACAAGTCTCTCATCCCAAGGACCCCGG---GGACCC 2454
|||||
Db 2475 CTGCCAAAGAGCTCGGGAGGACAAGTCTCTCATCCCGGACCCCGGGGAGGACCC 2534
|||||
QY 2455 TCTGTCTTGCAACACAGCGTCTCTCAGCCCTCTCCCAACCAAGGTGGAGCGCGGACGACAGS 2514
|||||
```

```
Db 2535 TCTGTCTTGCAACACAGCGTCTCAGCCTCTCCAAACAGGTGGAGCAGCGGACGACAGG 2594
|||||
QY 2515 CCCCTTCAGCTCGGTGGCTGGCAGCTGGGGCGCTGCTCCCGAGCTCGCGCAGTGGC 2574
|||||
Db 2595 CCCCTTCAGCTCGGTGGCTGGCAGCTGGGGCGCTGCTCCCGAGCTCGCGCAGTGGC 2654
|||||
QY 2575 CTGAGAAGCGGGCGGTGGAATGCCGGGCTCCGCCGGGACGCGACGCTCCCTGCTGT 2634
|||||
Db 2655 CTGAGAAGCGGGCGGTGGAATGCCGGGCTCCGCCGGGACGCGACGCTCCCTGCTGT 2714
|||||
QY 2635 GATGAGCCCATCGGCCGCTGGAGACACACAGCTTCGGGGGAGCCCTGCCACCTGGGAG 2694
|||||
Db 2715 GATGAGCCCATCGGCCGCTGGAGACACACAGCTTCGGGGGAGCCCTGCCACCTGGGAG 2774
|||||
QY 2695 CTGACGCGCTGTTCACCTGTCTCAAGAGCTTCGGCGCGGGGATTTTCAGAGCGCTCACTC 2754
|||||
Db 2775 CTGACGCGCTGTTCACCTGTCTCAAGAGCTTCGGCGCGGGGATTTTCAGAGCGCTCACTC 2834
|||||
QY 2755 AAGTGTGTGGGCCACGAGCGCGCTGTGGCCCGGGACAGTGCACCTTGACCCGCAAG 2814
|||||
Db 2835 AAGTGTGTGGGCCACGAGCGCGCTGTGGCCCGGGACAGTGCACCTTGACCCGCAAG 2894
|||||
QY 2815 CCCAGGAGCTGGACTTCTGCTCTGAGGCGCTGC 2850
|||||
Db 2895 CCCAGGAGCTGGACTTCTGCTCTGAGGCGCTGC 2930
|||||

RESULT 4
AP002986 157963 bp DNA linear PRI 08-MAR-2002
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:RP11-121M22,
DEFINITION complete sequence.
ACCESSION AP002986
VERSION AP002986.2 GI:19263031
KEYWORDS HTS.
SOURCE Homo sapiens DNA, clone:RP11-121M22.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in database (2000)
REFERENCE 2 (bases 1 to 157963)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Mar 7, 2002 this sequence version replaced gi:11559301.
FEATURES
Location/Qualifiers
1..157963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-121M22"

BASE COUNT 39742 a 36722 c 37247 g 44252 t
ORIGIN

Query Match 33.6%; Score 957.2; DB 9; Length 157963;
Best Local Similarity 98.2%; Pred. No. 3.8e-161;
Matches 968; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGCTTTTGTCTGGGCATCTCAACCTGGCTTCGCGCGGCGCAACCGCTTGGAGGCTTTGAG 60
|||||
Db 134115 ATGCTTCTGTGGGCATCTCAACCTGGCTTCGCGCGGCGCAACCGCTTGGAGGCTTCTGAG 134174
|||||
```

QY 61 CCAGAGCGGGAGGTAGTCTGTTCCCATCCGACTGACCGGACATTAAGCGCGCGCTAC 120
 Db 134175 CCAGAGCGGGAGGTAGTCTGTTCCCATCCGACTGACCGGACATTAAGCGCGCGCTAC 134234
 QY 121 TACTGGCGGGTCCGAGGACTCCGGGATACGGGACTCATTTTTCAGATCACAGCATTT 180
 Db 134235 TACTGGCGGGTCCGAGGACTCCGGGATACGGGACTCATTTTTCAGATCACAGCATTT 134294
 QY 181 CAGGAGGACTTTTACCTTACACCTGACGCGGATGCTCAGTTCTTGCTCCGCTCTCC 240
 Db 134295 CAGGAGGACTTTTACCTTACACCTGACGCGGATGCTCAGTTCTTGCTCCGCTCTCC 134354
 QY 241 ACTGAGCATCTGGCGGTCCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGCGACGC 300
 Db 134355 ACTGAGCATCTGGCGGTCCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGCGACGC 134414
 QY 301 TGTCTTATCTGGGACGTGACGCGGAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
 Db 134415 TGTCTTATCTGGGACGTGACGCGGAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 134474
 QY 361 GGGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCTGCCC 420
 Db 134475 GGGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCTGCCC 134534
 QY 421 AATGCTAGCGCGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 480
 Db 134535 AATGCTAGCGCGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 134594
 QY 481 GGTGTTCGGGCGGCTTCCGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 540
 Db 134595 GGTGTTCGGGCGGCTTCCGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 134654
 QY 541 AACCCCGCCATCTACGGGCGCTTGGACCTTACAAGCGCGGCGGCGGCTTCCGGGAG 600
 Db 134655 AACCCCGCCATCTACGGGCGCTTGGACCTTACAAGCGCGGCGGCGGCTTCCGGGAG 134714
 QY 601 AGTCGTAGCGCGGCGGAGGCTGCTGGGCGGCGGAGCGGAGCGGAGCGGAGCGGAG 660
 Db 134715 AGTCGTAGCGCGGCGGAGGCTGCTGGGCGGCGGAGCGGAGCGGAGCGGAGCGGAG 134774
 QY 661 GAGACGCTGTGTGTCGCGGAGGAGTCAATGGTCAAGTTCACGGGCGGAGCGGAGCGG 720
 Db 134775 GAGACGCTGTGTGTCGCGGAGGAGTCAATGGTCAAGTTCACGGGCGGAGCGGAGCGG 134834
 QY 721 TATCTGCTGACGCTGTGGAAGCGGCGGAGGAGTCTACCGGCGGAGCGGAGCGGAGCGG 780
 Db 134835 TATCTGCTGACGCTGTGGAAGCGGCGGAGGAGTCTACCGGCGGAGCGGAGCGGAGCGG 134894
 QY 781 CCATCAACATCTGTCGTCAGGTCGCTTCTTAGAGATCGTACGCGGCGGAGCGGAGCGG 840
 Db 134895 CCATCAACATCTGTCGTCAGGTCGCTTCTTAGAGATCGTACGCGGCGGAGCGGAGCGG 134954
 QY 841 GTACCGGCAATGCGGCGCTGAGCTGCGCAACTCTGTGCTGCGGAGGAGCGGAGCGG 900
 Db 134955 GTACCGGCAATGCGGCGCTGAGCTGCGCAACTCTGTGCTGCGGAGGAGCGGAGCGG 135014
 QY 901 AAGTGAAGTACAGACCGGAGTACTGGGAGACTGGCATCTCTTACCGAGGAGCGGAG 960
 Db 135015 AAGTGAAGTACAGACCGGAGTACTGGGAGACTGGCATCTCTTACCGAGGAGCGGAG 135074
 QY 961 CTGTGTGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 996
 Db 135075 AGTTGATGCGGCTGCTACCTTTGACACC 135100

RESULT 5
 AC025130/c
 LOCUS AC025130 172905 bp DNA linear HTG 24-AUG-2002
 DEFINITION Homo sapiens chromosome 11 clone RP11-211H6 map 11, WORKING DRAFT
 SEQUENCE, 16 unordered pieces.
 AC025130
 AC025130.2 GI:7387384
 HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS 1 (bases 1 to 172905)
 TITLE Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 JOURNAL Homo sapiens chromosome 11, clone RP11-211H6
 REFERENCE
 AUTHORS 2 (bases 1 to 172905)
 TITLE Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burdett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 172905)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burdett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 3, 2000 this sequence version replaced gi:7158941.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7808
 Center clone name: 211_H6
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads

SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS 1 (bases 1 to 172905)
 TITLE Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 JOURNAL Homo sapiens chromosome 11, clone RP11-211H6
 REFERENCE
 AUTHORS 2 (bases 1 to 172905)
 TITLE Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burdett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 172905)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burdett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 3, 2000 this sequence version replaced gi:7158941.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7808
 Center clone name: 211_H6
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
Consensus quality: 161899 bases at least Q40
Consensus quality: 167402 bases at least Q30
Consensus quality: 169801 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 171405; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 21: contig of 21 bp in length
22 121: gap of 100 bp
122 1343: contig of 1222 bp in length
1344 1443: gap of 100 bp
1444 2636: contig of 1193 bp in length
2637 2736: gap of 100 bp
2737 4052: contig of 1316 bp in length
4053 4152: gap of 100 bp
4153 8044: contig of 3892 bp in length
8045 8144: gap of 100 bp
8145 14949: contig of 6805 bp in length
14950 15049: gap of 100 bp
15050 23535: contig of 8486 bp in length
23536 23635: gap of 100 bp
23636 30261: contig of 6626 bp in length
30262 30361: gap of 100 bp
30362 40941: contig of 10580 bp in length
40942 41041: gap of 100 bp
41042 54893: contig of 13852 bp in length
54894 54993: gap of 100 bp
54994 68888: contig of 13895 bp in length
68889 68988: gap of 100 bp
68989 83702: contig of 14714 bp in length
83703 83802: gap of 100 bp
83803 98393: contig of 14591 bp in length
98394 98493: gap of 100 bp
98494 117356: contig of 18863 bp in length
117357 117456: gap of 100 bp
117457 140589: contig of 23133 bp in length
140590 140689: gap of 100 bp
140690 172905: contig of 32216 bp in length.

Location/Qualifiers
1. .172905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-211H6"
/clone_lib="RPC1-11 Human Male BAC"
1. .21
/note="assembly_fragment"
clone_end:T7
vector_side:right
122. .1343
/note="assembly_fragment"
1444. .2636
/note="assembly_fragment"
2737. .4052
/note="assembly_fragment"
4153. .8044
/note="assembly_fragment"
8145. .14949
/note="assembly_fragment"
15050. .23535
/note="assembly_fragment"
23636. .30261

/note="assembly_fragment"
clone_end:SP6
vector_side:right
30362. .40941
/note="assembly_fragment"
41042. .54893
/note="assembly_fragment"
54994. .68888
/note="assembly_fragment"
68989. .83702
/note="assembly_fragment"
83803. .98393
/note="assembly_fragment"
98494. .117356
/note="assembly_fragment"
117457. .140589
/note="assembly_fragment"
140690. .172905
/note="assembly_fragment"
BASE COUNT 44360 a 41333 c 39516 g 46194 t 1502 others
ORIGIN
Query Match 33.6%; Score 957.2; DB 2; Length 172905;
Best Local Similarity 98.2%; Pred. No. 3.8e-161; Indels 0; Gaps 0;
Matches 968; Conservative 0; Mismatches 18;
Qy 1 ATGCTTTTGTGGCATCCTAACCTTGCTTTCGCCGGCGAACCGCTGAGGCTTTGAG 60
Db 79583 ATGCTTCTGTGGCATCCTAACCTTGCTTTCGCCGGCGAACCGCTGAGGCTTCTGAG 79524
Qy 61 CCAGAGCGGGAGGTAGTGTTCCTCCATCGACTGGACCCGGACATTAACGGCGCGCTAC 120
Db 79523 CCAGAGCGGGAGGTAGTGTTCCTCCATCGACTGGACCCGGACATTAACGGCGCGCTAC 79464
Qy 121 TACTGGCGGGGTCCCGAGGACTCCCGGGATCCCGGGATCAGGGACTCATTTTCAGATCACAGCAATTT 180
Db 79463 TACTGGCGGGGTCCCGAGGACTCCCGGGATCAGGGACTCATTTTCAGATCACAGCAATTT 79404
Qy 181 CAGGAGGACTTTTACCTACACCTGACGGCGGATGTCTAGTCTTGGCTCCCGCTTCTCC 240
Db 79403 CAGGAGGACTTTTACCTACACCTGACGGCGGATGTCTAGTCTTGGCTCCCGCTTCTCC 79344
Qy 241 ACTGAGCATCTGGCGTCTCCCTCCAGGGGTCCAGGGGTCTTCAGACTGCGAGCGC 300
Db 79343 ACTGAGCATCTGGCGTCTCCCTCCAGGGGTCCAGGGGTCTTCAGACTGCGAGCGC 79284
Qy 301 TGCCTTCTATTTCTGGGACGTGAACGCCGCGGACTCGTTCGCTGCTGTGTAGCCTGTGC 360
Db 79283 TGCCTTCTATTTCTGGGACGTGAACGCCGCGGACTCGTTCGCTGCTGTGTAGCCTGTGC 79224
Qy 361 GGGGGGCTCCGGGAGGCTTTGGCTACCGAGCGCGAGTAGTATGTCATTTAGCCGCTGCC 420
Db 79223 GGGGGGCTCCGGGAGGCTTTGGCTACCGAGCGCGAGTAGTATGTCATTTAGCCGCTGCC 79164
Qy 421 AATGCTAGCGCGCGCGCGCAGCGCAACAGCGCGCGCACACTTCTCCAGCGCGCGG 480
Db 79163 AATGCTAGCGCGCGCGCGCAGCGCAACAGCGCGCGCACACTTCTCCAGCGCGCGG 79104
Qy 481 GGTGTTCCGGCGGGGCTTCCGGAGACCCCACTCTCGCTCGGGGGTGGCTTCGGGGTGG 540
Db 79103 GGTGTTCCGGCGGGGCTTCCGGAGACCCCACTCTCGCTCGGGGGTGGCTTCGGGGTGG 79044
Qy 541 AACCCCGCATCTCTAGGGGCTTGGACCTTACAGCGCGCGCGCGCGGCTTTCGGGGG 600
Db 79043 AACCCCGCATCTCTAGGGGCTTGGACCTTACAGCGCGCGCGCGCGGCTTTCGGGGG 78984
Qy 601 AGTCGTAGCGCGCGAGGTCTGGGGCGCGCAAGCGCTTTCGTCTATATCCCGGGTACGTG 660
Db 78983 AGTCGTAGCGCGCGAGGTCTGGGGCGCGCAAGCGCTTTCGTCTATATCCCGGGTACGTG 78924
Qy 661 GAGACGCTGGTGGTCCGGACGAGTCAATGTCAAGTTCACGCGCGCGGACCTTGAACAT 720
Db 78923 GAGACGCTGGTGGTCCGGACGAGTCAATGTCAAGTTCACGCGCGCGGACCTTGAACAT 78864

QY 721 TATCTGCTGACGCTGCTGCAACGGCGCGGACTTACCGCCATCCAGCATCCTCAAC 780
 |||||
 Db 7863 TATCTGCTGACGCTGCTGCAACGGCGCGGACTTACCGCCATCCAGCATCCTCAAC 78804
 |||||
 QY 781 CCATCAACATCGTTGTGGTCAAGTGCTGCTTCTTAGAGATCGTGACGCCGCGCAAG 840
 |||||
 Db 78803 CCCATCAACATCGTTGTGGTCAAGTGCTGCTTCTTAGAGATCGTGACGCCGCGCAAG 78744
 |||||
 QY 841 GTACCCGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAGCTGAAC 900
 |||||
 Db 78743 GTACCCGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAGCTGAAC 78684
 |||||
 QY 901 AAGTGTAGTGACAGACCCCGGAGTACTGGACACTGCCATCCTCTTACCAGGAGGAC 960
 |||||
 Db 78683 AAGTGTAGTGACAGACCCCGGAGTACTGGACACTGCCATCCTCTTACCAGGAGGAG 78624
 |||||
 QY 961 CTGTGTGAGGCCACCACTGTGACAC 986
 |||||
 Db 78623 AGTTGATCTGCGCTCACTTTTGACCC 78598
 |||||

RESULT 6

AC023429 170682 bp DNA linear HTG 10-APR-2001
 LOCUS Homo sapiens chromosome 11 clone RP11-121M22, WORKING DRAFT
 DEFINITION SEQUENCE, 7 unordered pieces.
 AC023429
 VERSION AC023429.19 GI:13569974
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 170682)
 AUTHORS Abola, A.P., Bruno, D., Conn, L., Della Rosa, M., Faulkner, D.,
 Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,
 Mao, J., Komp, C., Kottler, S., Lam, B., Marathe, R., Miranda, M.,
 Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D.,
 Southwick, A.M., Webb, C., Wilhelmy, J., Yu, S. and Davis, R.W.
 Unpublished

JOURNAL 2 (bases 1 to 170682)
 REFERENCE Bruno, D., Conn, L., Della Rosa, M., Faulkner, D., Federspiel, N.,
 Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,
 Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J.,
 Yu, S. and Davis, R.W.
 Direct Submission
 TITLE Submitted (14-FEB-2000) DNA Sequencing and Technology Center,
 JOURNAL Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA

COMMENT On Apr 10, 2001 this sequence version replaced gi:13562078.
 ----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center

Center code: SDSTDC
 Web site: <http://sequence-www.stanford.edu/group/human/>
 Contact: hum-info@sequence.stanford.edu
 ----- Project Information
 Center project name: 837
 Center clone name: RP11-121M22
 ----- Summary Statistics

Sequencing Vector: M13mp18; X02513
 Chemistry: Dye-primer; 12% of reads
 Assembly: Dye-terminator Big Dye; 86% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 165770 bases at least Q40
 Consensus quality: 167255 bases at least Q30
 Consensus quality: 167918 bases at least Q20
 Insert size: 172423; agarose-fp
 Insert size: 170082; sum-of-ctnigs
 Quality coverage: 9.0x in Q20 bases; agarose-fp
 Quality coverage: 9.1x in Q20 bases; sum-of-ctnigs.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1856: contig of 1856 bp in length
 * 1857 1956: gap of unknown length
 * 1957 7065: contig of 5109 bp in length.
 * 7066 7165: gap of unknown length
 * 7166 21089: contig of 13924 bp in length.
 * 21090 21189: gap of unknown length
 * 21190 34145: contig of 12956 bp in length
 * 34146 34245: gap of unknown length
 * 34246 48562: contig of 14317 bp in length
 * 48563 48662: gap of unknown length
 * 48663 109179: contig of 60517 bp in length
 * 109180 109279: gap of unknown length
 * 109280 170682: contig of 61403 bp in length.

FEATURES

Source

1. 170682
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone_lib="RP11-121M22"
 /clone_lib="RPC1 human BAC library 11"

1. 1856
 /note="assembly_name:Contig25"

1957. 7065
 /note="assembly_name:Contig26"

7166. 21089
 /note="assembly_name:Contig27"

21190. 34145
 /note="assembly_name:Contig28"

34246. 48562
 /note="assembly_name:Contig29"

clone_end:SP6
 48663. 109179
 /note="assembly_name:Contig30"

109280. 170682
 /note="assembly_name:Contig31"

clone_end:T7

BASE COUNT 43947 a 39732 c 39514 g 46884 t 605 others
 ORIGIN

Query Match

Best Local Similarity 31.3%; Score 893; DB 2: Length 170682;
 Matches 948; Conservative 0; Mismatches 36; Indels 4; Gaps 3;

QY 1 ATGCTTTTCTGGGCGATCCTAACCTGCTTTTCGCGGGCGAAACGCTGGAGGCTTTTCAG 60
 |||||
 Db 98460 ATGCTTTCTGCTGGCGATCCTAACCTGCTTTTCGCGGGCGAAACGCTGGAGGCTTTTCAG 98519
 |||||
 QY 61 CCAG-AGCGGGAGGTA-GTCGTTCCCATCCGACTGGACCCCGGACATTAACGCCGCCCT 118
 |||||
 Db 98520 CCAGCAGCGGGAGGTA-GTCGTTCCCATCCGACTGGACCCCGGACATTAACGCCCTTCGCT 98579
 |||||
 QY 119 ACTACTGGCGGGGTCGCCGAGGACTCCGGGGATCCAGGACTCATTTTTCAGATCAAGCAT 178
 |||||
 Db 98580 TTATCTGGCGGGTACC--GAGAGCCCGGGGTTCAGNACTCATTTTTCAGATCAAGCAT 98637
 |||||
 QY 179 TTCAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTTGGCTCTGCCTTCT 238
 |||||
 Db 98638 TTCAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTTGGCTCTGCCTTCT 98697
 |||||
 QY 239 CCACCTAGCATCTGGGCGTCCCTCCAGGGGCTACCGGGGGCTCTTTCAGACCTGGCAC 298
 |||||
 Db 98698 CCACCTAGCATCTGGGCGTCCCTCCAGGGGCTACCGGGGGCTCTTTCAGACCTGGCAC 98757
 |||||
 QY 299 GCTGCTTCTATTCTGGGACGTGAACCGCGGAGCTCGTTCGCTGCTGTGAGCCTGT 358
 |||||
 Db 98758 GCTGCTTCTATTCTGGGACGTGAACCGCGGAGCTCGTTCGCTGCTGTGAGCCTGT 98819
 |||||

```
QY 359 GCGGGGGCTCCCGGAGCCTTTGGCTACCGGCGCCGAGGATGTGTCATTAGCCGCTGC 418
|||||
Db 98818 GCGGGGGCTCCCGGAGCCTTTGGCTACCGGCGCCGAGGATGTGTCATTAGCCGCTGC 98877
QY 419 CCAATGCTAGCGCGCGCGGCGGCGGACGGAACAGCGGCGGCGACACCTTCTCCAGCGCC 478
|||||
Db 98878 CCAATGCTAGCGCGCGGCGGCGGACGGAACAGCGGCGGCGACACCTTCTCCAGCGCC 98937
QY 479 GGGGTGTTCGGGCGGCGCTTCGGGAGACCCACCTCTCGCTCGCGGGTGGCTCGGGCT 538
|||||
Db 98938 GGGGTGTTCGGGCGGCGCTTCGGGAGACCCACCTCTCGCTCGCGGGTGGCTCGGGCT 98997
QY 539 GGAACCCCGCATCTTACGCGGCGCTTGAGCCCTTACAAAGCGCGGCGGCGGCTTCGGGG 598
|||||
Db 98998 GGAACCCCGCATCTTACGCGGCGCTTGAGCCCTTACAAAGCGCGGCGGCGGCTTCGGGG 99057
QY 599 AGAGTCGTAGCGCGCGGCGAGCTGTGGGCGGCGGCGGCGGCGGCTTCGTCATCCGCGGTAGC 658
|||||
Db 99058 AGAGTCGTAGCGCGCGGCGAGCTGTGGGCGGCGGCGGCGGCGGCTTCGTCATCCGCGGTAGC 99117
QY 659 TGGAGAGCTGTGTGTCGGGCGGCGAGCTCAATGTCGAAGTTCACGCGGCGGCGGCGGCTTGAAC 718
|||||
Db 99118 TGGAGAGCTGTGTGTCGGGCGGCGAGCTCAATGTCGAAGTTCACGCGGCGGCGGCGGCTTGAAC 99177
QY 719 ATTATCTGTGAGCGCTGTGCGGAACGCGCGGCGGCGGCTTACCGCCATCCAGCATCCTCA 778
|||||
Db 99178 ATTATCTGTGAGCGCTGTGCGGAACGCGCGGCGGCGGCTTACCGCCATCCAGCATCCTCA 99237
QY 779 ACCCATCATACATCTGTGTGTCGAAGTGTCTTCTTAGAGATCGTGTACTCCGGGCCCA 838
|||||
Db 99238 ACCCATCATACATCTGTGTGTCGAAGTGTCTTCTTAGAGATCGTGTACTCCGGGCCCA 99297
QY 839 AGCTACCGGCAATCGGCGGCGCTGACGCTCGGCAACTTCTGTGCTGGCAGAGAAGCTGA 898
|||||
Db 99298 AGCTACCGGCAATCGGCGGCGCTGACGCTCGGCAACTTCTGTGCTGGCAGAGAAGCTGA 99357
QY 899 ACAAGTGTGACAGACGCGGCGGAGTGTGAGACTGCTTCTTACGAGTCTTCCAGCGGCG 958
|||||
Db 99358 ACAAGTGTGACAGACGCGGCGGAGTGTGAGACTGCTTCTTACGAGTCTTCCAGCGGCG 99417
QY 959 ACCTGTGTGGAGCCACCATCTGTGACAC 986
|||
Db 99418 TGAGTGTGTCGCTGCTACTTTCACCC 99445

RESULT 7
AC101990/c 182656 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone RP24-371J2, WORKING DRAFT SEQUENCE, 16 unordered
DEFINITION pieces.
AC101990
VERSION AC101990.2 GI:22381363
KEYWORDS HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 182656)
Birken,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP24-371J2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182656)
Birken,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgaltier,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,C., Horton,L., Horton,L., Hulme,W., Iliiev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182656)
Birken,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgaltier,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenka,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zemek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17060766.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17833
Center clone name: 371J.2
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 178030 bases at least Q40
Consensus quality: 179964 bases at least Q30
Consensus quality: 180647 bases at least Q20
Insert size: 17200; agarose-fp
Insert size: 181156; sum-of-contents
Quality coverage: 8.1 in Q20 bases; agarose-fp
Quality coverage: 7.7 in Q20 bases; sum-of-contents
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 361: contig of 361 bp in length
* 362 461: gap of 100 bp
* 462 1538: contig of 1077 bp in length
* 1539 1638: gap of 100 bp
```

```

* 1639 3397: contig of 1759 bp in length
* 3398 3497: gap of 100 bp
* 3498 4947: contig of 1450 bp in length
* 4948 5047: gap of 100 bp
* 5048 6893: contig of 1846 bp in length
* 6894 6993: gap of 100 bp
* 6994 10395: contig of 3402 bp in length
* 10396 10495: gap of 100 bp
* 10496 12584: contig of 2089 bp in length
* 12585 12684: gap of 100 bp
* 12685 14977: contig of 2293 bp in length
* 14978 15077: gap of 100 bp
* 15078 18181: contig of 3104 bp in length
* 18182 18281: gap of 100 bp
* 18282 22654: contig of 4373 bp in length
* 22655 22754: gap of 100 bp
* 22755 31261: contig of 8507 bp in length
* 31262 31361: gap of 100 bp
* 31362 43587: contig of 12226 bp in length
* 43588 43687: gap of 100 bp
* 43688 57612: contig of 13925 bp in length
* 57613 57712: gap of 100 bp
* 57713 77467: contig of 19755 bp in length
* 77468 77567: gap of 100 bp
* 77568 108724: contig of 31157 bp in length
* 108725 108824: gap of 100 bp
* 108825 182656: contig of 73832 bp in length.

```

FEATURES

```

source
1..182656
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RPCI-24 Male Mouse BAC"
1..361
/note="assembly_fragment"
462..1538
/note="assembly_fragment"
1639..3397
/note="assembly_fragment"
3498..4947
/note="assembly_fragment"
5048..6893
/note="assembly_fragment"
6994..10395
/note="assembly_fragment"
10496..12584
/note="assembly_fragment"
12685..14977
/note="assembly_fragment"
15078..18181
/note="assembly_fragment"
18282..22654
/note="assembly_fragment"
22755..31261
/note="assembly_fragment"
31362..43587
/note="assembly_fragment"
43688..57612
/note="assembly_fragment"
57713..77467
/note="assembly_fragment"
77568..108724
/note="assembly_fragment"
108825..182656
/note="assembly_fragment"
BASE COUNT 51468 a 41160 c 39770 g 48758 t 1500 others
ORIGIN

```

Query Match 27.3%; Score 779.8; DB 2; Length 182656;
 Best Local Similarity 87.1%; Pred. No. 1.6e-129;
 Matches 856; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 1 ATGCTTTTGTGGGCATCTTAACCTGGCTTCGGCGGCGGAACCGCTGGAGGCTTTGAG 60

```

Db 39782 ATGCTTCTACTGGGCATCTCAATCTCGGCTTCGGCTGGCGACCTGCTGGCGAGCTCCGAG 39723
QY 61 CCAGAGCGGAGGTAGTCTGTTCCCATCCGACTGGACCGGACATTAACGGCCGCGCTAC 120
Db 39722 CCAGAGTGGAGGTGGTCTGTTCCAAATCCGACGGGACCGGACATCAATGGCCGCGCTAC 39663
QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGACTATTTTCAGATCACACATTT 180
Db 39662 TACAGAGGGGTACGAGGAGACTCCGGGGATCAGGCTCTCATTTTCAGATCACACATTTT 39603
QY 181 CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCTTCCTCC 240
Db 39602 CAGCAGGACTTTTATCTACACCTGACCGCAGATGCCAGTCTCTGGCTCCCGCTTTGCT 39543
QY 241 ACTGAGACTCTGGGGTCCCGCTCCAGGGGCTCAGGGGGCTCTTCAGAGCTCGGAGCG 300
Db 39542 ACTGAGTATCTAGGTGTCCTCCGCTGCAGAGGCTCACTGGCAGCTCTAGACCTGAGCG 39483
QY 301 TGTCTTCTATTCTGGGACGTGAACCGCGAGCTCGTTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db 39482 TGTCTTCTATTCTGGGTATGTGAACCGCGAGCTCCTTTGCTGCTGCTGCTGCTGCTGCTG 39423
QY 361 GGGGGCTCCCGGAGCTTTGGCTACCGAGCGCGGAGTATGTCTATAGCCGCTGCC 420
Db 39422 GGGGGTCTCCGCGAGCTTTGGCTACCGAGGTGGGAGTATGTCTATAGCTCTGCTGCC 39363
QY 421 AATGCTAGCGCGCGCGCGCGAGCAACGCCAGGGGGCACACTTCTCCAGCGCGG 480
Db 39362 AACACCGCGCGCGAGCGCGAGCTCACAGCCAGGGGCGACACTTCTCCAGCGCGG 39303
QY 481 GGTGTTCCGGGGCGGCTTCGGGAGACCCACTCTCGCTGGGGGTGGCTCGCGGCTG 540
Db 39302 GGTGCTCTGTAGGCGCTTCGGAGACCCCACTCTCTGTGGGGGTGGCTCGCTGCTG 39243
QY 541 AACCCTGCTTACGGGCGCTTGAGCCCTTACAAGCGCGCGCGGCGGCTTCCGAGCTG 600
Db 39242 AACCCCGCATCTCTGAGGGCTCTGGACCTTATAAGCCACCGCGGAGCGGCGGAG 39183
QY 601 AGTCGTAGCGCGCGAGCTCTGGGGCGCGCAAGCGTTTCGTGTCTATCCCGGGTACGTG 660
Db 39182 AGCCACAACCGCGCGAGGTCTGGGGCGCGCAAGCGCTTCGTGTCTATACCACGCTAG 39123
QY 661 GAGAGCTGTGTCGCGGAGGAGTCAATGGTCAAGTTCACAGCGCGCGGAGCTGGAACAT 720
Db 39122 GAGACACTGGTGGCGGAGGAGTCAATGGTCAAGTTCACAGCGCGGAGTTCGAACT 39063
QY 721 TATCTGTGACGCTGCTGGCAACCGCGCGGAGTCTTACCGCCATCCAGCATCTCAAC 780
Db 39062 TATCTACTGAGCTGCTGGCCACCGCGCGGAGTCTTACCGCCACCCAGCATCTCAAC 39003
QY 781 CCCATCAACATCGTTGCTCAAGGTGCTGCTTCTTAGAGATGCTGACTCGGGGCGGCAAG 840
Db 39002 CCTATCAACATCGTTGCTCAAGGTGCTGCTTCTTAGAGATGCTGACTCGGGGCGGCAAG 38943
QY 841 GTCACGGCAATGGGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAGCAAGCTGAAC 900
Db 38942 GTCACAGGCAACGGCGCTGACTCTGGCAACTTCTGTGCTGGCAGAGCAAGCTGAAC 38883
QY 901 AAAGTGTGACAGACAGCCCGAGTACTGGGACACTGCCATCTCTTTCACAGGCGAGGAC 960
Db 38882 AAAGTGTGACGATAAGCACCCCTGAGTACTGGGACAGCAGCATCTCTTCCACCAGACAG 38823
QY 961 CTGTGTGGAGCCACCATCTGTGA 983
Db 38822 AGTTCACTGTGCACCTCTTTGGA 38800

```

RESULT 8
 AP003459
 LOCUS
 DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-211H5,
 complete sequence.


```

Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 178764)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 178764)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZFU
Center clone name: CH230-254N12
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124573 bases at least Q40
Consensus quality: 131850 bases at least Q30
Consensus quality: 137700 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1037: contig of 1037 bp in length
* 1038
* 1138
* 2250: contig of 1113 bp in length
* 2251
* 2350: gap of unknown length
* 2351
* 3579: contig of 1229 bp in length
* 3580
* 3679: gap of unknown length
* 3680
* 4874: contig of 1195 bp in length
* 4875
* 4975: gap of unknown length
* 4976
* 6525: contig of 1551 bp in length
* 6526
* 6625: gap of unknown length
* 6626
* 7905: contig of 1280 bp in length
* 7906
* 8006
* 9724: contig of 1719 bp in length
* 9725
* 9824: gap of unknown length
* 9825
* 11442: contig of 1618 bp in length
* 11443
* 11542: gap of unknown length
* 11543
* 12900: contig of 1358 bp in length
* 12901
* 13000: gap of unknown length
* 13001
* 14150: contig of 1150 bp in length
* 14151
* 14250: gap of unknown length
* 14251
* 15565: contig of 1315 bp in length
* 15566
* 15665: gap of unknown length
* 15666
* 17475: contig of 1810 bp in length
* 17476
* 17575: gap of unknown length
* 17576
* 18690: contig of 1115 bp in length
* 18691
* 18790: gap of unknown length
* 18791
* 20269: contig of 1479 bp in length
* 20270
* 20369: gap of unknown length
* 21927: contig of 1558 bp in length
* 21928
* 22027: gap of unknown length
* 22028
* 23998: contig of 1971 bp in length
* 23999
* 24098: gap of unknown length

```

```

24099 25679: contig of 1581 bp in length
* 25680
* 25779: gap of unknown length
* 27428: contig of 1649 bp in length
* 27429
* 27528: gap of unknown length
* 27529
* 29777: gap of unknown length
* 29778
* 30377: contig of 1100 bp in length
* 30378
* 30977: gap of unknown length
* 30978
* 32831: contig of 1854 bp in length.
* 32832
* 32931: gap of unknown length
* 32932
* 35419: contig of 2488 bp in length
* 35420
* 35519: gap of unknown length
* 35520
* 37393: contig of 1874 bp in length
* 37394
* 37493: gap of unknown length
* 37494
* 39835: contig of 2342 bp in length
* 39836
* 39935: gap of unknown length
* 39936
* 42789: contig of 2854 bp in length
* 42790
* 42889: gap of unknown length
* 42890
* 46646: contig of 3757 bp in length
* 46647
* 46746: gap of unknown length
* 46747
* 49188: contig of 2442 bp in length
* 49189
* 49288: gap of unknown length
* 49289
* 53805: contig of 4517 bp in length
* 53806
* 53905: gap of unknown length
* 53906
* 56202: contig of 2297 bp in length
* 56203
* 56302: gap of unknown length
* 56303
* 58719: contig of 2417 bp in length
* 58720
* 58819: gap of unknown length
* 58820
* 62904: contig of 4085 bp in length
* 62905
* 63004: gap of unknown length
* 63005
* 65932: contig of 2928 bp in length
* 65933
* 66032: gap of unknown length
* 66033
* 70120: contig of 4088 bp in length.
* 70121
* 70220: gap of unknown length
* 70221
* 73258: contig of 3038 bp in length
* 73259
* 73358: gap of unknown length
* 73359
* 77466: contig of 4108 bp in length
* 77467
* 77566: gap of unknown length
* 77567
* 80545: contig of 2979 bp in length
* 80546
* 80645: gap of unknown length
* 80646
* 85987: contig of 5342 bp in length
* 85988
* 86087: gap of unknown length
* 86088
* 90908: contig of 4821 bp in length
* 90909
* 91008: gap of unknown length
* 91009
* 97918: contig of 6910 bp in length
* 97919
* 98018: gap of unknown length
* 98019
* 103502: contig of 5484 bp in length.
* 103503
* 103602: gap of unknown length
* 103603
* 112570: contig of 8968 bp in length
* 112571
* 112670: gap of unknown length
* 112671
* 117924: contig of 5254 bp in length
* 117925
* 118024: gap of unknown length
* 118025
* 125019: contig of 6995 bp in length.
* 125020
* 125119: gap of unknown length
* 125120
* 134087: contig of 8968 bp in length
* 134088
* 134187: gap of unknown length
* 134188
* 143851: contig of 9664 bp in length
* 143951: gap of unknown length
* 143952
* 151699: contig of 7748 bp in length
* 151700
* 151799: gap of unknown length
* 158826: contig of 7027 bp in length.
* 158927
* 158926: gap of unknown length
* 158927
* 168522: contig of 9596 bp in length
* 168523
* 168622: gap of unknown length
* 168623
* 178764: contig of 10142 bp in length.
FEATURES
Location/Qualifiers
1..178764
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-254N12"
47487 a 39119 c 38897 g 45423 t 7838 others
BASE COUNT
ORIGIN

```

Query Match 27.2%; Score 775.2; DB 10; Length 3927;
Best Local Similarity 85.4%; Pred. NO. 2.1e-128;
Matches 864; Conservative 0; Mismatches 148; Indels 0; Gaps 0

QY 1842 CAAGTGCAGCTCATCTGCCGAGCCAATGGCACTGGCTACTTCTATGCTGCTGGCAACCA 1901

```
Db 13 CAAGTGTAACTAATGTCGCCAGCAATGGGACTGGCTACTTCTATGTAAGTACACCTAA 72
QY 1902 GGTGTGTGACGGCAGCTGCTCTCTCTGACTCCACCTCGTGTGTCTCAAGCAAGTG 1961
Db 73 GGTGTGTGAGTGTAGCTGTGTCTCTGACTCCACCTCGTGTGTGTCTCAAGCAAGTG 132
QY 1962 CATCAAGGCTGGCTGTGATGGGAACTTGGGCTCCAAAGAGAGATTCGACAAAGTGGGGT 2021
Db 133 CATCAAGGCTGGCTGGCAGCGGAATCTGGCTCCAAAGAGAAATTTGACAAATGTGGTG 192
QY 2022 GTGTGGGGGACACAAATAGAGCTCAAGAGAGTGTGAGTCTTTCACCAAGCCCATGCA 2081
Db 193 GTGTGGGTGGAGACAAATAGAGCTTAAGAGGGTGACAGGACTCTTTCACCAAGCCTATGCA 252
QY 2082 TGGCTACAATTTTCGTGTGGCATCCCGCAGGCGCTCAAGCATCGACATCGCCACGG 2141
Db 253 CGGCTACAATTTTGTAGTGGCCATCCCTGTGCGGCTCCAGCTTATATATCCGTACGG 312
QY 2142 CGGTTACAAAGGCTGATCGGGGATGACAACTACTCGGCTCTGAAGAACCAAGCA 2201
Db 313 TGGTTACAAGGGCTCATTTGAGATGACAACTACTCGGCTTTGAAGAACAGCCAAAGCAA 372
QY 2202 GTACTGTCTCAAGCGCATTTCTGTGTGTCGGGTGGAGCGGACCTGTGTGTGAAGG 2261
Db 373 ATACTGTCTCATGGGCACTTTGTGGTATCCGCTGTAGAGCGGGACCTGTGTAAAG 432
QY 2262 CAGTCTGCTGCGGTACAGCGGACGCGGACAGCGGTGGAGAGCTGTGAGGCTTCCCGGCC 2321
Db 433 CAGTGTGCTACGCTATAGTGGCACTGTACTGCAGTGGAGAGCTGCAGGCTTCTCGACC 492
QY 2322 CATCTGTGAGCGCTGACGTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGGGT 2381
Db 493 CATCTGTGAGCGCACTACCGTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGGGT 552
QY 2382 CCGCTACTCTCTTCTATCTGCCAAAGAGCTCGGGAGCAAGTCTCTCTATCCCAAGGA 2441
Db 553 GCGTTATCTCTTACTCTGCCAAAGAACTCGGGAGCAAGTCTCTCTATCCCAAGGA 612
QY 2442 CCCCCGGGACCCCTCTCTTTGCACAAAGCGTCTCTCAGCCTCTCCAAACAGGTGGAGCA 2501
Db 613 CCGCTGAGGTTCCCTCTGTTGGCAACAGTGTCTCAGCCTTTCCAAACAGTAGAGCA 672
QY 2502 GCGGAGCAGACAGCCCTCGACGTGGTGGTGGCTGGGAGCTGGGGCCGTGTCGCGGAG 2561
Db 673 GCGAGATAACAGACCTCTCGCGGCTGGGTGGCAGGAGCTGGGGACCTTGTCTCGGTGAG 732
QY 2562 CTGCGGCACTGGCTGACAGAGCGGCGGTGGACTGCGCGGGCTCGCGGCGGAGCGCAC 2621
Db 733 CTGTGGCAGTGGCTACAGAGCGAGCAGTGGACTGCGCGGACTCCCCAGGCGAGCAAG 792
QY 2622 GGTCCCTGCCTGTGATGACAGCCATCGGCCCTGGGAGACACAAGCCTCGGGGAGCCCTG 2681
Db 793 AGCCTCTGCTGTGATGTAGACCATCGGCCATTGAGAGCGAGCCTGTGGGAAACCCCTG 852
QY 2682 CCCCACCTGGAGCTCAGCGCTGCTCACCCTGCTCCAGAGCTGGGGCCGGGATTTCA 2741
Db 853 CCGAATTTGGAGCTCGCAACTGTGCGCCCTGTCTTAAAGCTGTGGCGGTGGATTA 912
QY 2742 GAGCGCTCACTCAAGTGTGTGGGCCACGAGGCGCGGCTGCTGGCCCGGAGCACAGTGCA 2801
Db 913 GAGCGCTCCGCTCAAGTGTGTGGGCCACGAGGCGAGGCTGCTGGCTCGGGACCAAGTGA 972
QY 2802 CTTTCCACCGCAAGCCCGAGAGCTGGACTTCTGCGTCTGAGCGCGGTGCTGA 2853
Db 973 CTTACGCGGCAAGCCCGAGGAATTAGACTTCTGCGTCTTGAGACCCCTGCTGA 1024

RESULT 11
E55282 LOCUS 2670 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel metalloprotease and gene of the same.
ACCESSION E55282
VERSION E55282.1 GI:18629795
```

```
KEYWORDS JP 2001008687-A/18.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2670)
AUTHORS Yamaji.N., Nishimura.K. and Sasamata.M.
TITLE Novel metalloprotease and gene of the same
JOURNAL Patent: JP 2001008687-A 18 16-JAN-2001;
YAMANOUCHI PHARMACEUT CO LTD
COMMENT OS Homo sapiens (human)
PN JP 2001008687-A/18
PD 16-JAN-2001
PF 25-JUN-1999 JP 1999180973
PR NOBORU YAMAJI,KOICHI NISHIMURA,MIHO SASAMATA
PC C12N15/09,C07K16/40,C12N1/15,C12N1/19,C12N1/21,C12N5/10, PC
C12N9/64,C12Q1/37,
PC C12N15/00,C12N5/00
CC
FH Key Location/Qualifiers
FT source 1..2670
FT /organism='Homo sapiens (human)'.
FEATURES
Source
Location/Qualifiers
1..2670
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 511 a 853 c 867 g 439 t
ORIGIN
Query Match 23.3%; Score 666; DB 6; Length 2670;
Best Local Similarity 59.2%; Pred. No. 6.6e-109;
Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;
QY 157 CTCATTTTTCAGATCACAGCATTTTCAGGAGGACTTTTACCTACACTGAGCGCCGATGCT 216
Db 157 CTCGCGCTCCACTGTCCGCTTCGGCAAGGCTTCGCTGCTGCGCTCCGCGCCGACACG 216
QY 217 CAGTTCTTGGCTCCCGCTTCTCCACTGAGCATCTGGCGTCCCGCTCCCGCTCCAGGGCTCAC 276
Db 217 AGCTTCTTGGCGCCGAGTTCAGATCGAGCCCTCGGGGGCTCCGGCGGCG---ACC 273
QY 277 GGGGCTCTTTCAGACTCGGACGCTGCTCTTATTCTGGGAGCTGAACCGCAGCGGAC 336
Db 274 GGGGCGAGCGGGGGCTGCGCGCTGCTCTTCTCCGCGACCGCTGAATGGGAGCGCCGAG 333
QY 337 TCGTTCGCTGCTGAGCGCTGTCGGGGGGCTCCCGGAGCCTTTGGCTACCGAGCGCC 396
Db 334 TCGTGGCGGCGGTGAGCGCTGTCGGCGGCTGAGCGGCTCCTTCTGCTGGACGGCGAG 393
QY 397 GAGTATGTCTATTAGCCCGCTGCCCAATGTAGCGCGCGCGCGCGCGCACGCAACAGCCAG 456
Db 394 GAGTTACACTCCAGCCGACGCGCGCGGGGCTCCCTGGCTCAGCCCGCACCGCTGCGAG 453
QY 457 GCGGACACCTTCTCCAGGCGCGGGGTGTTCGGGGGGCGCTTTCGGAGACCCCACTCT 516
Db 454 CGCTGGGCTCCCGCGGAGCGCCCGCTCCCGCGAGGACCGCGAGTGGAGGTGGAGAGC 513
QY 517 CGCTCGGGGGTG-GCCTCGGGCTGGAACCCCGCATCTACGGGCGCTGGACCTTTACAA 575
Db 514 GGAGAGGTTCAGAGCAGGAGAGAGACCAACAGAGGAGCAGCGAGAGAGAGAGCA 573
QY 576 GCGCGCGCGGGCGGCTTCGGGGAGAGTCTGAGCGGGGCGCAGG-----TCTGGCGCGCC 630
Db 574 GAAGAGAGGAGAGAGGCGCTAGCGAGCGCGCCACCGCCCTGGGGGCCACGAGTAGGACC 633
QY 631 AAGCGTTTCTGTCTATCCCGGGTACGTGGAGAGCGCTGTGGTGTGCGGAGAGTCAATG 690
Db 634 AAGCGTTTGTCTCTGAGCGCGCTTCGTGAAGAGCGCTGCTGTGGCGCGATCGCTCCATG 693
QY 691 GTCAAGTTCCAGCGCGCGGACCTGGAACATTATCTGCTGACGCTCTGCTGGCAACGCGCG 750
Db 694 GCTGCTTCTACGGGGCGGACCTCGAGAACCACATCTCTGAGCTTAATGTC---GTGACAGCC 753
```

QY 751 CGACTTACCCGATCCAGCATCTCTCAACCCCAATCAATCGTGTGTGTCATAGGTGCTG 810
D 754 CGAATCTACAGACGACCCAGCATCAAGAAATTCATCAACCTGATGTGTGTAAAGTCTG 813
QY 811 CTTCTTAGAGATCTGACTCGGCGCCCAAGTCAACCGCAATCGGCGCTGAGCGTCCG 870
D 814 ATCGTAGAGATGAATAATGGGCGCCAGAGGTGTGCAACAATGGGGGCTTACACCGT 873
QY 871 AACTTCTGTGCTGGCAGAGAACTGAACAAAGTGAAGTGAACAGCAACCCGAGTACTGG 930
D 874 AACTTCTGCACTGGCAGCGCTTCAACCAACCGCAGCAACCCGAGCACTAC 933
QY 931 GACATGCGCATCTCTTACACAGGCAAGCTGTGTGG---AGCCACCACTGTGACACC 987
D 934 GACACGGCCATCTCTCTCACAGACAGAACTTCTGTGGCAGGAGGGGCTGTGTGACACC 993
QY 988 CTGGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAGAACTGCTGTGATTCAG 1047
D 994 CTGGGTGGCAGACATCTGGGACCATTTGTGACCCCAACAAAGCTGCTCCGTGATCGAG 1053
QY 1048 GACATGGGCTTCCATCAGCTTCCACCACTGCCCACGAGCTGGGCCACGCTGTTCACATG 1107
D 1054 GATGAGGGCTCCAGCGGCGCCACACCTTGCCCATGAAGTGAAGGACGCTCTCAGCATG 1113
QY 1108 CCCATGACAAATGTGAAGTCTGTGAGAGGTGTGTGGAGGTCCGAGCAACACCATG 1167
D 1114 CCCACGACGACTCAAGCCCTGCACACGCTCTTCGGGCCATGGGCAAGCACCAGCTG 1173
QY 1168 ATGTCCCGGACCTTATCCAGATGACGCTGCCCACGAGCTGGGCCACGCTGTGAGTGC 1227
D 1174 ATGGCAGCTGTGCTGCTTCCACTGAACACGAGCTGCTGCTGGTCCCTCGAGCGCATG 1233
QY 1228 ATATCACCGACTTCTGGACAGCGGCGGAGTACTGCTCTCTGGACCAACCCAGCAAG 1287
D 1234 TATCTCAGAGCTTCTGGACGCGGCGGACGAGAGTGTCTCTGGATGCCCTGTGCG 1293
QY 1288 CCCATCTCCCTGCGGAGGATCTGCGGCG---GCCAGCTACACCTTGAGCGCAGCAG 1341
D 1294 GCGCTGCGCCCTCCACAGGCTTCCGGGCGCATGCGCTGTACAGCTGGACCAAGCAG 1353
QY 1342 TCGAGATGGCTTTTGGCGTGGCTTCAAGCCCTGCTCTTACA-----TGCAGTAC 1392
D 1354 TCGAGGAGATCTTTGGCGCGATTTCCGCCACTGCTGCTCAGGAGCTC 1413
QY 1393 TGCACCAAGCTGTGGT---CACCGGAAGGCCAAGGAGAGATGTGTGCCAGACCCG 1449
D 1414 TCGCCCAAGCTTGGTGCCACACTGATGGGCTGAGCCCTGTGCCACAGAAATGGC 1473
QY 1450 CACTTCCCTGGCGGATGGCAGCAGCTGTGGGAGGCAAGCTCTGCTCAAGGGGCG 1509
D 1474 AGCTTGGCTGGCTGAGCGGACCCGTGGGCGCTGGGACCTCTGCTCAGAGGCGAGC 1533
QY 1510 TG-----CGTGAGAGACAACTCAACAAAGCAGAGGTGGATGTTCTTGGGCCAAA 1563
D 1534 TGTCTACTGAGGAGGAAGTGGAGAGCCCAAGCCCTGTAGATGAGGCTGGCACCG 1593
QY 1564 TGGATCCCTATGCGCCCTGCTGCGCAGATGTGTGGGCGGTGACGTGGCCAGAGG 1623
D 1594 TGGGAGCCCTGGGAGAAATGTTCTCGACCTGTGGAGGAGGTACAGTTTTCACACCGT 1653
QY 1624 CAGTGCACCAACCCCTGCGCAAGCGGGCAAGTACTGCGAGGAGGTGAGGTGAAA 1683
D 1654 GAGTGCAGAGGCCCGAGCTCAGATGGAGGAAGATACTGCTTGGGTGGAGAGCCAG 1713
QY 1684 TACCAGTCTGCAACCTGAGCCCTGCGCCAGCTCAGCCCTCGGAAAGAGCTTCCGGGAG 1743
D 1714 TACCAGTCTGCAACCTGAGCCCTGAGAGGAGGAGTGGGCTGAGAGGAGGAG 1767
QY 1744 GAGCAGTGTGAGGCTTTCACAGGCTACACCAAGCAGCAACCCGCTCACTCTCGCGCTG 1803
D 1768 CAGCAGTGTGAGAGATATATGCTTACATTTACACTGACATGGACGGGAATCT---CCTG 1824

QY 1804 GCATGGTCCCAAGTACTCCGGCGTCTCTCCCGGACAAAGTCAAGCTCATCTGCCGA 1863
D 1825 CAGTGGGTCCCAAGTATGCTGGGTGTCCCGCGGACCGCTGCAAGTGTCTGCCGA 1884
QY 1864 GCAATGGCAGTGGTACTTCTATGTGTGCGCACCACCAAGTGTGAGCGGACGCTGTGC 1923
D 1885 GCGCGGAGAGGAGGTCAAGTGTTCAGAGCCAGGTGATTTGATGGACCTGTGT 1944
QY 1924 TCTCTGACTCCACCTCCCTCTGTGTCAAGGCAATGCATCAAGGCTGTGTATGGG 1983
D 1945 GGGCCAGAAACACTGCCATCTGTGTCCGTGGCCAGTGTCAAGCGCGCTGTACCAT 2004
QY 1984 ACTTGGGCTCCAGAGAGATTCCAGAGTGTGGGTGTGTGGGGGAGACATAGAGC 2043
D 2005 GTGTGGACTCCCTCGGAAGCTGGACAAATCGGGGTGTGTGGGGGCAAGGCAACTCC 2064
QY 2044 TCAAGAAGGTGACTGGGACTCTTCAACAGCCATGCATGGCTACAATTTCTGTGGTGC 2103
D 2065 TCAAGAAGGTCTCCGGGTCCCTCACCCACCAATTTAGGCTACANTGACATTTGTACC 2124
QY 2104 ATCCCGCGAGGCGCTCAAGCATCGACATCCCGCGGCTTACAAAGGCTGTATCGGG 2163
D 2125 ATCCAGCTGTGTGCCACTAATATTGACGTGAAGCAGCGGAGCCACCGGTGTGCAGAAC 2184
QY 2164 GATGACACTACTGCTGCTGAAGACAGCCAGCAAGTACCTGCTCAACGGGCAATTC 2223
D 2185 GATGGAATACTGCTGGCGCTGAAGAGCGGTGATGGGAGTACCTGCTCAACGGCAACTG 2244
QY 2224 GTGTGTCCGGGTGGAGGGGACCTGTGTGTGAAGGCGAGTCTGCTCGGTGACAGCGC 2283
D 2245 GCAATCTCTGCATAGAGGAGGACATCTGTGTGAAGGGGACCATCTGAGTACAGCGGC 2304
QY 2284 ACGGACAGCGGTGAGAGCCTCGAGGCTTCCCGGCCATCTCTGGCGGCTGACCGTG 2343
D 2305 TCAATCGCACCTGAGGCGCTGAGAGCTTCCGGCGCTTCCAGAGCCTCTGACAGTG 2364
QY 2344 GAGGTCTCTCGT---GGGGAAGATGACACCGCCCGGCTGCTACTCTCTTATCTG 2400
D 2365 CAGTCTGACAGTCCCTGGCGAGGTCTCCCGGAGGAGTCTCCCGGCAAGTACAAATACCTCTTCTGT 2424
QY 2401 CCCAAGA 2408
D 2425 CCTAATGA 2432
RESULT 12
E58655
LOCUS
DEFINITION Novel metallic protease.
ACCESSION E58655
VERSION E58655.1 GI:18629877
KEYWORDS JP 2001017183-A/3.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2670)
AUTHORS Yamaji, N., Nishimura, K. and Sasamata, M.
TITLE Novel metallic protease
JOURNAL Patent: JP 2001017183-A 3 23-JAN-2001;
YAMANOUCHI PHARMACEUT CO LTD
COMMENT OS Homo sapiens (human)
PN JP 2001017183-A/3
PD 23-JAN-2001
PF 09-JUL-1999 JP 1999196584
PR NOBORU YAMAJI, KOICHI NISHIMURA, MIHO SASAMATA
PC C12N15/09, C07K16/40, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
C12N9/50, C12O1/37,
PC C12N15/00, C12N5/00
CC
FH Key Location/Qualifiers
FT source 1. .2670

FEATURES		FT	Location/Qualifiers		/organism='Homo sapiens (human)'	
source			1..2670			
BASE COUNT		511 a	853 c	867 g	439 t	
ORIGIN						
Query Match 23.3%; Score 666; DB 6; Length 2670;						
Best Local Similarity 59.2%; Pred. No. 6.6e-109;						
Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;						
Qy	157	CTCATTTTTCAGATCACAGCATTTTCAGGAGGACTTTTACCTACACCTGACCTGACGCGGATGCT	216			
Db	157	CTCGCGCTCCACCTGTCCGCTTCGGCAAGGCTTCGTGCTGGCCTGGCGCCGACGAC	216			
Qy	217	CAGTTCTTGGCTCCGCTCTTCGACTGAGCATCTGGGCTGCCCTCCAGGGCTCACC	276			
Db	217	AGCTTCTGGCGCCGAGTTCAAGATCGAGCGCTCGGGGCTCCGGCCGGCG--ACC	273			
Qy	277	GGGGGCTCTTCAGACCTTCGACGCTGCTTCTATTCTGGGACCTGAACGCGGACGCGGAC	336			
Db	274	GGGGGCGAGCGGGGCTGCGCGCTGCTTCTTCTCCGGCACCTGATGGGAGGCCGAG	333			
Qy	337	TCGTTCTGCTGTGTGAGCTGTGCGGGGGCTCCCGGAGCCTTTGGCTACCGAGCGCC	396			
Db	334	TCGCTGGCGGGGTACGCTGTGCGGGGGCTGAGCGGCTCCTTCTGCTGGACGCGGAG	393			
Qy	397	GAGTATGTCATTAGCCCGCTGCCAATGCTAGCGCGCGCGGCGGCGACGCGCAACAGCCAG	456			
Db	394	GAGTTCACCATCCAGCGCAGGCGCGGGGGCTCCCTGGCTCAGCGCGCACCGCTGCGAG	453			
Qy	457	GGCGCACACCTTCTCCAGCGCGGGGTGTTCGGGGCGGGCTTCCGGAGACCCCACTCT	516			
Db	454	CGCTGGGCTCCGCGGAGCGCCGCCCTCCCGCGAGGACCCGAGTGGAGGTGGAGAG	513			
Qy	517	CGCTCGCGGGTG--CCCTCGGGCTGGAACCCGCCATCTTACGGGCGCTTGGACCTTACAA	575			
Db	514	GGAGAGGCTCAGAGCGAGGAGAGAGACACCGAGGAGACGCGAGGAGGAGAGGCCAA	573			
Qy	576	GCCGCGCGGGCGGGCTTCGGGGAGAGTCTGTAGCGCGCGCAGG-----TCTGGCGCGCGC	630			
Db	574	GAAGAGGAGCGAGAAGCGCTAGCGAGCGCGCACCGCCCTGGGGGCCACGAGTAGGACC	633			
Qy	631	AAGCTTTCGTGTCTATCCCGCGTACGTGGAGACGCTGGTGTGCGGGACGAGTCAATG	690			
Db	634	AAGCGGTTTGTCTGAGGCGCGCTTCGTGAAGACGCTGCTGTGGTGGCGATGCTGCTCATG	693			
Qy	691	GTCAAGTTCCAGCGCGGACCTGGAACATTATCTGTGAGCTGTGGCAACGCGCGCG	750			
Db	694	GCTGCCCTTACGGGGCGGACCTGCAGAACCATCTGACCTTAATGTCGTGGCAGCC	753			
Qy	751	CGACTTACCGCATCCAGCATCTCAACCCCATCAACATCGTTGTGGTCAAGGTGCTG	810			
Db	754	CGAATCTTACAAGCACCCAGCATCAAGAATTCATCAACCTGATGTTGGTAAAGTCTG	813			
Qy	811	CTTCTTAGAGATCGTACTCCGGGCCAAGTCAACGGCAATGCGGCCCTGACGCTGCGC	870			
Db	814	ATCGTAGAAGATGAAAATATGGGCGCCAGAGGTGTCCGACAATGGGGGCTTACACTGCT	873			
Qy	871	AACCTTGTGCTGCGCAGAGAGTGAACAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG	930			
Db	874	AACCTTGTCAACTGGCAGCGGCGTTTCAACCGACCGCAGCGCCACCCAGACACTAC	933			
Qy	931	GACACTGCCATCTCTTACCGAGGAGGACCTGTGTGG-----AGCCACCACTGTGACACC	987			
Db	934	GACACGCCCATCTGCTTACAGACAGAACTTCTGTGGGCGAGGAGGGGCTGTGTGACACC	993			
Qy	988	CTGGGATCGCTGATGTGGGTACCATGTGTGACCCCAAGAGAGGCTGCTCTGTTCATTGAG	1047			
Db	994	CTGGGTGTGCACACATCGGGACCAATTTGTGACCCCAACAAAAGCTGCTCCGTGATCGAG	1053			
Qy	1048	GACGATGGGCTTCCATCAGCCCTTACCACCTGCCACGAGCTGGGCCACAGTGTTCACATG	1107			

Db	1054	GATGAGGGCTCCAGCGGCCACACCTCGGCCCATAGCTAGGACAGTCTCTAGCATG	1113		
Qy	1108	CCCATGACAAATGTGAAGTCTGTGAGGAGGTGTTTGGGAAGCTCCGAGCCAAACACATG	1167		
Db	1114	CCCCAGGAGCTCCAAAGCCCTGCACACGGCTCTTTGGGCCCATGGCAACACACATG	1173		
Qy	1168	ATGTCGCCGACCTCATCCAGATCGACGTCGCAACCCCTGGTCAAGCTGCACTGCTGCC	1227		
Db	1174	ATGGCACCGCTTCTGCTCCACCTGAACACAGACGCTGCCCTGGTCCCTGAGTCCATG	1233		
Qy	1228	ATCATCACCGACTTCTTGACAGCGGACGCTGCTCTCTGAGCAACCCAGCAAG	1287		
Db	1234	TATCTCACAGACTTCTTGACGCGGCGGACGAGACTGTCTCTGATGCTGCTGCTGCG	1293		
Qy	1288	CCATCTCCCTGCCGAGATCTGCCGGC-----GCCAGCTACACCTGAGCCAGAG	1341		
Db	1294	GCCCTGCCCTCCACAGACGCTCCCGGCCCATGAGCCCTGTACAGCTGACACGAG	1353		
Qy	1342	TCGAGCTGGCTTTTGGGTGGGCTCCAAAGCCCTGCTCTTACA-----TCAGTAC	1392		
Db	1354	TCAGCGAGATCTTTGGCGGATTTCCGCCACTGCCCAACACCTCTGCTCAGACGCTC	1413		
Qy	1393	TCACCAAGCTGTGTG---CACCGGGAAGCCAAAGGACAGATGCTGTGCCACACCGC	1449		
Db	1414	TCGCGCCAGCTTTGTGTGCCACACTGATGGGCTGAGCCCTGTGCCACACGAAGTGGC	1473		
Qy	1450	CACCTCCCTGGCGGATGCGACAGCTGTGGCAGGAGCAAGCTGTGCTCTAAAGGGGCC	1509		
Db	1474	AGCTTGCCCTGGGCTGACGCGCGCTGCGGGCTGGGCGCTGCTGTCAAGAGGACG	1533		
Qy	1510	TG-----CGTGGAGACACAACTCAACAGCACAGGCTGGATGTTCTTGGGCCAAA	1563		
Db	1534	TGTTCTACCTGAGGAGGAGTGGAGAGGCCAAAGCCCTGTGATGAGGCTGCGCACCG	1593		
Qy	1564	TGGATTCCTTAFGCCCCCTGTGCGGCACATGTGTGTGGGGCGTGCAGCTTGCAGGAGG	1623		
Db	1594	TGGGAGCCCTGGGAGAAATGTTCTCGGACCTGTGTGGAGGAGTACAGTTTTCACACCGT	1653		
Qy	1624	CAGTCCACCAACCCACCCCTGCCAACGGGGGAGTACTCGGAGGAGTACAGGTGAA	1683		
Db	1654	GAGTCAAGAGCCCGAGCCCTCAAGATGAGAGAAATGCTGCTGGTGGAGAGATCCAAAG	1713		
Qy	1684	TACCGATCTGCAACTGGAGCCCTGCCCGAGCTCAGCCTCCGGAAGAC---TTCGCGGAG	1743		
Db	1714	TACCATGTCATGCCACAGGAGAAATGCCCC-----CCTGACGGGAAAG---TTCAGGGAG	1767		
Qy	1744	GAGCAGTGTGAGGCTTCAACGGCTAACCCACAGACCAACCGGCTCCTCTCGCGGTG	1803		
Db	1768	CAGCAGTGTGAGAGTATAATGCTTACAAATACACTGACATGGACGGAAATCT---CCTG	1824		
Qy	1804	GCATGGTCCCAAGTACTCTCGGCGTGTCTCCCGGACAAAGTCAAGCTCATCTGCCGA	1863		
Db	1825	CAGTGGTCCCAAGTACTGTGGGTGTTCCCGGAGCCCTGCAAGTTCTTCTGCCGA	1884		
Qy	1864	GCAATGGCACTGGCTACTTCTGTGTGTCACCCCAAGTGTGGAGGTCACCTGTGTC	1923		
Db	1885	CCCGGGGAGGAGGAGTCAAAAGTGTTCGAGGCCAAGGTGATTGATGG---ACCTGTGT	1944		
Qy	1924	TCTCTGACTCCACTCGTCTGTGTCCAAGGAGTGCATCAAGGCTGGTGTGTGATGGG	1983		
Db	1945	GGGCCAGAAACACTTGGCCATCTGTCTCGTGGCAGTGTGTCAAGGCGGCTGTGACCAT	2004		
Qy	1984	AACCTGGCTCCAAAGAGAGATTCGACAAGTGTGGGTGTGTGGGGGAGACAATAAGAGC	2043		
Db	2005	GTGTGGACTCGCTCGGAAGCTGGCAAAATCCGGGGTGTGTGGGGGAAAGGCAATCC	2064		
Qy	2044	TGCAAGAGGTGACTGGACTTTTCCCAAGCCCATGCTGCTGCTGCTGCTGCTGCTGCTG	2103		
Db	2065	TGCAAGAGGTGCTCGGGTCTCTCACCCCAACCAATATGCTGCTGCTGCTGCTGCTGCTG	2124		
Qy	2104	ATCCCCGAGGCGCTCAGCATCGCCAGCGGCTTACAAAGGCGCTGATGCGG	2163		

Db 2125 ATCCAGCTGGTCCCACTAATATGACGTGAAGCAGCGGACCCACCGGCTGTCAGAAC 2184
Qy 2164 GATGACAACTACCTGGCTCTGAAGACAGCAGCAAGGCAAGTACCTGCTCAACGGGCAATTC 2223
Db 2185 GATGGGAACCTACCTGGCGCTGAAGAGCGGTGATGGCGAGTACCTGCTCAACGGCAACCTG 2244
Qy 2224 GTGCTGTCGGCGGTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGGCGGTACAGCGGC 2283
Db 2245 GCCATCTCTGCCATAGCAGGACATCTTGGTGAAGGGGACCATCTCTGAAGTACAGCGGC 2304
Qy 2284 ACGGGACACGGGTGAGAGCCTGAGGCTTCCCGGCCCATCTCTGAGCGCGCTGACCGTG 2343
Db 2305 TCCATGCCACCTGGAGCGCCGTCAGAGCTTCCGGCCCTTCCAGAGCCTCTGACAGTG 2364
Qy 2344 GAGTCTCTCCGT---GGGGAAGATGACACCGCCCGCGGTCCGCTACTCTCTCTATCTG 2400
Db 2365 CAGTCTCTGACAGTCCCTGGCGAGGTCTTCCCGCCCAAGTCAAAATAGACCTCTTGT 2424
Qy 2401 CCCAAAGA 2408
Db 2425 CCTAATGA 2432

RESULT 13
AF060153 3711 bp mRNA linear PRI 11-AUG-1999
LOCUS Homo sapiens METH2 protein (METH2) mRNA, complete cds.
DEFINITION AF060153
ACCESSION AF060153.1 GI:5725507
VERSION
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3711)
AUTHORS Vazquez,F., Hastings,G., Ortega,M.A., Lane,T.F., Oikemus,S.,
Lombardo,M. and Iruela-Arispe,M.L.
TITLE METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a
new family of proteins with angio-inhibitory activity
J. Biol. Chem. 274 (33), 23349-23357 (1999)
MEDLINE 99367466
PUBMED 10438512
REFERENCE 2 (bases 1 to 3711)
AUTHORS Vazquez,F., Hastings,G., Ortega,M.-A., Lane,T.F., Lombardo,M.,
Oikemus,S. and Iruela-Arispe,M.L.
TITLE Direct Submission
JOURNAL Submitted (16-APR-1998) Pathology, Beth Israel Deaconess Medical
Center, 99, Brookline Avenue, Boston, MA 02214, USA
FEATURES
source 1..3711
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3711
/gene="METH2"
706..3378
/gene="METH2"
/note="extracellular protein with metalloprotease and TSP
domains"
/codon_start=1
/product="METH2 protein"
/protein_id="A048081.1"
/db_xref="GI:5725508"
/translation="MFPAPAAPRWLFLPFLLLLLLLPLARGAPAPRPAAGQASLVVP
TRLPVSAGELALHLSAFGKGVFLRLAPDSDFLPAPEFKIERLGGSGRATGEGRLKGF
FSTVNGPELSLAVALSCLRGLSGFLDGEFTIQPGAGSLAQPHRLQWRGPGAR
PLPRGPEWETEGEQRQGHQDESESEAEASEPPPLIGATSRTRKRVSE
ARVETLLVADASMAAFYGADLQNHILTLMSVAARIYKHPSTKNSINLWVKVLIVED
EKGPVSDNGGILTLPNFCNWRFRNPDRHPEHYDTAILTRONFCQEGCLDTLG
VADIGTCDPNKSVCVIEGLQAHTLAHELGHVLSMHPDSDSKPTRLFGPWGRHHV
MAPLFLHNLTLFPSPCSAMYLFTELIDGGHGDCLLDAPGAALPLTGLPGRWALDOLD
QOCRLQIFEEVERPKVSDQVCAHLDGAEPLCHTKNGSLPWADPTCPGPHLC
SEGSLPFEERPKVSDQVCAHLDGAEPLCHTKNGSLPWADPTCPGPHLC
LGRRAVQSCHTECPDGKSFREQCEKYNAFYTMDGNLLQWPKYAGVSPDRIC

KLFCRARGSEFKVFEAKVIDGTLGPELTAICVRGCYKAGCDHVVDSRKLKCGV
CGKNSCRVSLQETDITVYNDIVIPAGATNIDVKORSHGVQNDYVLLAKLTAD
GOYLLNGNLAISETQIDILVKYSGIATLERLSRPLQVPLTVOLLVLPGEV
PPKVKYTFVFPNDVDFSMOSSSKERATTNIIQPLHQAQWLVGWSSECSTCGAGWRR
TVCEKDPGSOASATCNKALPDKAPKCESQLCLP"
BASE COUNT 733 a 1126 c 1201 g 646 t 5 others
ORIGIN
Query Match 23.3%; Score 666; DB 9; Length 3711;
Best Local Similarity 59.2%; Pred. No. 6.2e-109;
Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;
Qy 157 CTCATTTTTCAGATCAGACATTTTCAGGAGACATTTTACCTACACCTGACGCCCGATGCT 216
Db 865 CTCGCGCTCCACCTCTCCGCTTCGGCAAGGGCTTCGTGTTGCGCGCTGGCCCGCGAGAC 924
Qy 217 CAGTTCTTTGGCTCCCGCTTCTCCACTGACGATCTGGCGCTGCCCTCCAGGGGCTCAC 276
Db 925 AGCTTCCTGGCGCCCGAGTTCAAGATCAGCGCTCGGGGCTCGGCGCGGGG---ACC 981
Qy 277 GGGGCTCTTCAGACCTCGACGCTGCTTCTATTCTTGGGGACGTGAACGCGCAGCGCGGAC 336
Db 982 GGGGCGAGCGGGGCTCGCGGCTGTTTTTTTCCGCGACCGTGAATGGGAGCGCGAG 1041
Qy 337 TCGTTCGCTGCTGTCAGCCTGTGGCGGGGCTCCGCGGAGCCTTTGGCTACGAGCGGCC 396
Db 1042 TCGCTGGCGGGCTGAGCCTGTGCGCGGGCTGAGCGGCTCCTCTCCTCTGAGCGCGAG 1101
Qy 397 GAGTATGTCTATTAGCCCTGCCCCAATGCTAGCGCGCGCGGGGCGACGCGCAACAGCCAG 456
Db 1102 GAGTTCACCATCCAGCGCGAGCGCGGGGCTCCCTGGCTCAGCGCGCACCGCTGCAG 1161
Qy 457 GCGCACACCTTCTTCAGCGCGGGGCTGTTCCGGCGGGGCTTCGGGAGACCCCACTCT 516
Db 1162 CGCTGGGCTCCCGCGGAGCGCCCGCTCCGCGGAGGACCCGAGTGGGAGGTGGAGACG 1221
Qy 517 CGCTGGCGGGT-GCCTCGGGCTGGAACCCCGCCATCTACGGGCGCTGGACCC---TACA 575
Db 1222 GGAGAGGTTCAGAGCGAGGAGAGAGACACACAGGAGGACGAGGAGAGAGAGACCA 1281
Qy 576 GCGCGGGCGGGGCTTCGGGGAGAGTGTAGCGCGCGCAGG-----TCGGGCGCGCC 630
Db 1282 GAAGAGGAGCGAAGAGGCGCTAGCGAGCGCCCGCCCGCTGGGGCGCCACGAGTAGGAC 1341
Qy 631 AGCGTTTCTGCTCTATCCCGCGGTACGTGGAGAGCGTGTGTCGCGSAGAGTCAATG 690
Db 1342 AAGCGGTTGTCTGTAGCGCGCTTCGTGGAGAGCGTGTGTCGCGCGATGCTGCTCATG 1401
Qy 691 GTCAAGTTCACGCGCGGACCTGGAACATTATCTGTGCTGACGCTGCTGGCAACGCGCG 750
Db 1402 GCTGCTTCTAGCGGCGGACCTGCAGAACACATCTGACGTTAATGCTGTGCGACGC 1461
Qy 751 CGACTCTACGGCCATCCCAAGCATCTCAACCCCATACATCGTTGTGGTCAAGTGTG 810
Db 1462 CGAATCTACAAGCACCCAGCATCAAGAAATTCATCAACCTGATGTGTGTTAAAGTGTG 1521
Qy 811 CTTCTTAGAGATCGTACTCCCGGGCCCGAGGTCACCGCAATGCGCGCTGACGCTGCG 870
Db 1522 ATCGTAGAAGATGAAATAATGGGGCCCGAGAGGTGTCGACAAATGGGGGCTTACACTG 1581
Qy 871 AACTTCTGTGCTCGCAGAGAAGCTGAACAAAGTGAAGTGAACAAGCACCCCGAGTACTGG 930
Db 1582 AACTTCTGCACTGGCAGCGGCGTTTCAACAGCGCCAGCGCCGCCAGAGCACTAC 1641
Qy 931 GACATGCGCATCTCTTACAGGAGGAGGACCTGTGTGG---AGCCACCACTGTGACACC 987
Db 1642 GACACGCCCATCTCTGCTACACAGAGAACTTCTGTGGGCGAGGAGGCGCTGTGTGACACC 1701
Qy 988 CTGGGCGATCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1047
Db 1702 CTGGGCTGGCAGACATCGGAGACATTTGTGACCCCAACAAAGCTGCTCGTGATCGAG 1761
Qy 1048 GAGGATGGGCTTCCATCATGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1107

Db	2833	ATCCAGCTGGTCCACTAATATTGACGTGAAGCAGCGAGCCACCCGGGTGTCAGAAC	2892
QY	2164	GATGACAACCTACCTGGCTCTGAAGAACACGCCAAGCAAGTACCTGCTCAACGGGCATTTTC	2223
Db	2893	GATGGGAACCTACCTGGCGCTGAAGACGGCTGATGGGCAAGTACCTGCTCAACGGCAACCTG	2952
QY	2224	GTGCTGTGCGCGGTGGAGCGGACCTGGTGGTGAAGGCGAGTCTGCTCGGTACAGCGGC	2283
Db	2953	GCCATCTCTGCCATAGACGACGACATCTTGGTGAAGGGGACCATCCTGAGTACAGCGGC	3012
QY	2284	ACGGGACAGCGGTGGAGAGCTTGCAGGCTTCCCGGCCCATCTCTGGAGCGCTGACCGTG	2343
Db	3013	TCCATCGCCACCCCTGGAGGCGCTTGCAGAGCTTTCGGCGCCCTTGCAGAGCCCTCTGACAGTG	3072
QY	2344	GAGTGCTCTTCGGT---GGGGGAAGATGACACCGCCCGGGTCCGGTACTCTCTTCTATCTG	2400
Db	3073	CAGCTCTTGACAGTCCCTGGCGAGGTCTTCCCCCAAAAGTCAAAATACACCTTCCTTGTT	3132
QY	2401	CCCAAGA	2408
Db	3133	CCTAATGA	3140

RESULT 14	
D67076	
LOCUS	4180 bp mRNA linear F0D 02-FEB-1999
DEFINITION	Mouse mRNA for secretory protein containing thrombospondin motifs, complete cds.
ACCESSION	D67076
VERSION	D67076.1 GI:1813339
KEYWORDS	secretory protein containing thrombospondin motifs; ADAMTS-1.
SOURCE	Mus musculus colon adenocarcinoma cell_line:murine colon 26 cDNA to mRNA.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; 1 (sites)
REFERENCE	Kuno,K., Kanada,N., Nakashima,E., Fujiki,F., Ichimura,F. and Matsushima,K. Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene J Biol Chem. 272 (1), 556-562 (1997)
AUTHORS	97150761
TITLE	2 (bases 1 to 4180) Kuno,K., Kanada,N., Nakashima,E., Fujiki,F., Ichimura,F. and Matsushima,K. Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene Unpublished
JOURNAL	3 (bases 1 to 4180) Kuno,K.
MEDLINE	Direct Submission
AUTHORS	Submitted (28-SEP-1995) Kouji Kuno, Cancer Research Institute, Kanazawa University, Pharmacology; Takara-machi 13-1, Kanazawa, Ishikawa 920, Japan (Tel:0762-62-8151(ex.5454), Fax:0762-60-7704)
JOURNAL	Ishikawa 920, Japan (Tel:0762-62-8151(ex.5454), Fax:0762-60-7704) Location/Qualifiers
FEATURES	1. .4180 source /organism="Mus musculus" /db xref="taxon:10090" /cell_line="murine colon 26" /cell_type="colon adenocarcinoma" 1. .4180 /gene="ADAMTS-1" 441..3296 /gene="ADAMTS-1" /note="putative" /codon_start=1 /product="secretory protein containing thrombospondin motifs"
gene	
CDS	

/translation="MGDVORARSRGSLSAHMLLLLLASITMILLCARGAHORPIEDEE
 ELVLSLRAPGHDSTTRRLRDLAFQQLHLKLOPDSGLFAPGFTLQVGRSPSEAO
 HLDPTGDLAHCYFSGTVNGDPGSAALSLCEVRGAFYLOGEEFFIOPAPGATERLA
 PAVPEESARQFHLILRRRRSGGAKGVMDDETLPSTDSRPESONTRNWPVRDP
 TPQDAGTSKPGSIRKRFVSPRYVETMLVADMDHFGSLKHLKLLTLFVAAARE
 YKHSIRNSISLVVVKILVLYEOKGPEVTSNAALTILNFCWSKQKHSNDRDPEHY
 DPAILFTQDLGSHGSDTLHMADVGTVCDPFSRSCSVYEDDLQAFTTAHELGHVEN
 MPHDAKHCSLNGVSDHMLMSLSLSDHSPWSPCSAYMVTSLDNHGHCEIMDK
 PNPILKPLSDPLFLYDANROQFTFGEEKHCPDAASTCTLMCTGTSGLLVCOTK
 HPFWADGTSCGEGKVCWGKCNKMDKFATPVHSGMPWGPWGDSCRITCGGVQVOT
 MRECDNPVKGNGKYCEGRVRYRSCNIEDCPDNNKGTFRFEEQCEANEFKSAFSGNE
 PIVEWTPKYAGVSPDKCKLCEAKIGCYFVYLOPKYVDGTPCPSDSTSVYCVQCVK
 AGCDRIIDSKKFKGCVGGNGSTCKKWSGIYVSTRPGYHDIIVIPAGATNIEVKHR
 NORGRRNGSELAIRADGTYLILNGFNLTSLLEODLYTKGLVLRVSGSSAALERISF
 SPKLEPLTVQVLMHALRPKIKYTFMKKTESFNAIPFSEWVIEWEGSCSTCS
 GQWRVVOQDINGHPASECAKEVPKASTPCADLPCHPMQVGDWSPCSKTRGKGYK
 RFLKCVSHDGGVLSNESCDPLKPKHYIDFCTLTQCS"

BASE COUNT 1078 a 1051 c 1146 g 905 t
 ORIGIN

Query Match 22.3%; Score 636; DB 10; Length 4180;
 Best Local Similarity 60.7%; Pred. No. 1.3e-103;
 Matches 1098; Conservative 0; Mismatches 700; Indels 12; Gaps 3;

QY 631 AAGCGTTTCGTCTATCCGCGGTACGTGGAGACGCTGGTGGTGGCGGACAGTCAATG 690
 DB 1143 AAGCGATTTGTTCAGCCCCCGTTATGTGGAAACATGCTCGTAGCCAGCCATG 1202
 QY 691 GTCAAGTTCCAGCGCGGACCTGGAAACATTATCTGCTGACGCTGTGCGCAACGGCGG 750
 DB 1203 GCGGACTCCAGCGGACGGTCTAAGACATTACCTTCTAACCTGTCTCGGTGGCAGC 1262
 QY 751 CGACTCTACGGGATCCCAAGCATCTCAACCCCATCAACATCGTTGTGTCGAAGTGTG 810
 DB 1263 AGGTTTTTAAAGCATCCCAAGCATTAGGAATTCAATTAGCTGTGCTGGTGAAGATCT 1322
 QY 811 CTTCCTTAGAGATCGTACTCCGCGGCGCAAGTACCGCAATGCGGCGCTGACGCTGGC 870
 DB 1323 GTCATATAGAGGACGAGGACGAGGATTTACTTCCATGACGCTCTCAACCTCGG 1382
 QY 871 AACTTCTGTGCTTGGCAGAGAGCTGAAACAAAGTGAAGTGAACAGCACCCCGAGTACTGG 930
 DB 1383 AATTTCTGACGTGGCAGAAACACACACACCCAGTGGACCGGATCCAGAGCACAT 1442
 QY 931 GACACTGGCATCTCTTACAGGAGGACCTGTGTGGAGCCACACCTGTGACACCTG 990
 DB 1443 GACACTGCAATCTGTTCACAGAGAGGATTTATGTGGCTCCACACGCTGTGACACTCTC 1502
 QY 991 GGATCGCTGATGTGGGTACCATGTGACCCCAAGAGAGAGTCTCTCTCATTCAGGAC 1050
 DB 1503 GGATGGCAGATGTTGGAAACCGTATGTGACCCACAGAGGCTGCTCAGTCATAGAAT 1562
 QY 1051 GATGGCTTTCCATGACGCTTACCCATGCCCCAGAGCTGGGCGACGCTGTCAACATGCCC 1110
 DB 1563 GATGGTTTCAAGCGGCTTACCCAGCCCATGAATTTGGGCCATGTGTTTAAACATGCG 1622
 QY 1111 CATGACATGTGAAGTCTGTGAGGAGTGTGTTGGAGAGCTCCGAGCCCAACACATGATG 1170
 DB 1623 CACGATGATGCTAAGCACTGTGCCAGCTTGAATGTTGAGTGGCGATTCATCTGATG 1682
 QY 1171 TCCCGGACCTCATCAGATGACCGTGCACACCGCTGTGACGCTGAGTCTGCTGCCATC 1230
 DB 1683 GCTCGATGCTCTCAGCTTAGACCATAGCCAGCCCTGGTCACTTGGTGGCTGCTACATG 1742
 QY 1231 ATCACCAGCTTCTGAGACGCGGACGCTGACTGCTGCTGAGCAGCAACCCAGCAAGCCC 1290
 DB 1743 GTCAGCTCTCTTAGATATGACACGGGAAATGTTGATGGACAGACCCAGATCCA 1802
 QY 1291 ATCTCCTCCGAGATGTCGCGGCGCGAGCTACACCCCTGAGCCAGCAGTGGAGGTG 1350
 DB 1803 ATCAAGCTCCCTCTGATGCTTCCCGGTACCTTGTACGATGTCACACCGCCAGTGTG 1862
 QY 1351 GCTTTTGGGCTGCCAAGCCCTGTCTTACATGCAAGTAC---TGACCAAGCTGTGG 1407

DB 1863 ACATTCGGAGAGGAATCCAGCACTGCGCTGATGACGACGACATGTACTACCTGTGG 1922
 QY 1408 TGCACCGGAAGGCAAGGGACAGATGTGTGGCCAGACCGCCACTTCCCTGSGGCGGAT 1467
 DB 1923 TGCACCTGCACCTCGGTGGCTTACTGTGTGCGCAAAACAAACACTTCCCTTGGGACAT 1982
 QY 1468 GGCACACGCTGTGGCGAGGCAAGCTCTGCTCTAAAGGGGCTCGCTGGAGAGACAAAC 1527
 DB 1983 GGCACACGCTGTGGAGAGGGAAGTGTGTGTCAGTGCAGTGCAGTGCAGTGCAGTGCAG 2042
 QY 1528 CTCAACAGCACAG-----GGTGGATGTTCTTGGGCGCAAAATGGGATCCCTATGCCCC 1581
 DB 2043 ATGAAGCATTTTGTACTCTCTGTTTTCATGGAAGCTGGGACCATGGGACGCTGSGGAGAC 2102
 QY 1582 TGCTTCGCGCACATGTGTGGGGGCGTCAGCTGCCAGGAGGACATGCACAAACCCCAAC 1641
 DB 2103 TGCTCAAGACCTGTGGTGGTGGAGTTCATATACATATGAGATGTGACAAACCCAGTC 2162
 QY 1642 CCTGCCAACGGGGCAAGTACTGCGAGGGAGTGAAGGTGAATACCGATCCTGCAACCTG 1701
 DB 2163 CCAAGAACGAGGGAAGTACTGTGAAGGCAACAGAGTCCGCTACAGGTCTCTGTAAACATC 2222
 QY 1702 GAGCCTGCCCCACCTCAGCTCCGGAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTC 1761
 DB 2223 GAGGACTGTCCAGACAATAA---CGGAAACGTTTCAGAGAGGAGCAGTGGGAGGCGCAC 2279
 QY 1762 AACGGCTACAAACACAGCACCACCGCTCCTCTCGCCGTGGCATGGTGGTCCCAAGTAC 1821
 DB 2280 AATGATTTTCCAAAGCTTCTTTTGGGAATGAGCCACTGTAGAGTGGACACCCAGTAC 2339
 QY 1822 TCCGGGTGTCTCCCGGGGACAGTGAAGTCAATGTCGCGAGCAATGGCAGTGGGTAC 1881
 DB 2340 GCGGCGCTCTCGCCAAAGGACAGGTGCAAGTCAACCTGTGAAGCCAAAGGCAATGGGTAC 2399
 QY 1882 TTCTATGTGTGGCCACCAAGGTGTGGAGGAGCAGCTGTGCTCTCTGACTCCACCTCC 1941
 DB 2400 TTTTTCGTTTACAGCCCAAGGTTGTAGTGGCACTCCCTGTAGTCCAGACTCTACCTCT 2459
 QY 1942 GTCTGTGTCCAAGCAAGTGCATCAAGGCTGGCTGTGTGATGGGAACCTGGGCTCCAAAG 2001
 DB 2460 GTCTGTGTGAAGGCAAGTGTGTGAAGCTGGCTGTGATCGCATCATAGACTCCAAAG 2519
 QY 2002 AGATTCACAAAGTGTGGGTGTGTGGGGGAGACATTAAGAGCTGCGAAGAGGTGACTGGA 2061
 DB 2520 AAGTTTGAATAGTGTGGCTTGTGGAGGAACGGTTCCACATGCAAGAAAGATGTCAGGA 2579
 QY 2062 CTCTTCACCAAGCCATGCTGCTGCTACAAATTTGCTGTGTGGCTATCCCGCAGCGCTCA 2121
 DB 2580 ATAGTCACTAGTACAAACCTGGGTATCATGACATTTGTCACAAATTCCTGCTGGAGCCACC 2639
 QY 2122 AGCATTCGACATCCCGCCAGCGGTTTACAAAGGGCTGATCGGGGATGACAAATACCTGGCT 2181
 DB 2640 AACATTTGAAGTGAACATCGAATCAAGGGGTGCCAGAAACAAATGCGACCTTTCCTGGCT 2699
 QY 2182 CTGAAGAACACCAAGCAAGTACTCTCTCAAGGGGCAATTTGCTGTGTGGGCTGGAG 2241
 DB 2700 ATTAGACCGCTGT 2759
 QY 2242 CGGACCTGT 2301
 DB 2760 CAAGACCTTCACTTCAAGAGT 2819
 QY 2302 AGCCTGACGCTTCCCGGCGCATCTCTGGAGCGCTGACCGTGGAGGTCTCTCTCCGCTGGG 2361
 DB 2820 AGATTCGCGAGCTTGTAGTCCACTCAAGAACCTTAAACATCCAGGTTCTTATGCTAGGC 2879
 QY 2362 AAGATGACACGCGCGGCTCCGCTACTCTCTTATCTGCGCCAAAGAGCCTCGGAGGAC 2421
 DB 2880 CATGCTCTCGGACCAAAATTAATTCACCTACTTTATGAAGAAGAGACAGAGTCAATTC 2939
 QY 2422 AAGTCTCTC 2431

Db	740	TTT	CACGCTC	CAGAACG	TGGGG	CGCAAA	TCCGGT	TCCGAG	ACGCGG	TTCCGGA	ACCAG	799
QY	292	CTG	CAGCGCT	GTCTT	ATTTCT	TGTTGGG	ACGTGA	AGCCG	AGCGAG	CTGTTT	CGTCTGCT	351
Db	800	CTG	CGGCACT	GTCTT	CTACT	CCGAC	CGTGA	TATGG	GATCC	AGCTCGG	CTGCC	859
QY	352	AGC	CTGTG	CGGGGG	GTCTCC	CGAGG	CTTTGG	TCTAC	CGAGG	CGCGAG	TATGTCATT	411
Db	860	AGC	CTGTG	CGAGG	CGGTG	CGCGG	CTTCT	ACTCT	GCTGGG	AGGCGTA	TTTCATCC	919
QY	412	CCG	TGCCCC	AAATG	TATGCGG	CGC	CGGCG	GCAC	CGCCAA	CAGCC	AGGCGCAC	470
Db	920	CCG	TGCCCC	CGCCAG	CGAG	CGCTT	CGCAC	CGCGC	CGCCAG	GGGAGA	CGCGCGCA	979
QY	471	CC	AGCGC	CGGGGT	TTTCCG	GGCGG	CGCTT	CCG	AGAC	CCCACT	CTCTCG	530
Db	980	CCA	CTAG	CTTCA	CTCTT	CGCGCG	GAATCG	GCAG	GGGAG	CTGCGG	CGCACG	103
QY	531	CTG	GGGTGA	AACCCG	CCATCT	TACGGG	CCCTTGA	CACTT	CAAG	CGCGG	CGCGGG	590
Db	1040	GTG	TGGAC	CAG	GAGCCCG	CGCTG	GGAA	AGCG	AGAC	CGAAG	CAGGAC	109
QY	591	CTT	CGGGG	AG	AGTGT	AG	CGGCG	CGAG	GTG	GGGCG	-----	627
Db	1100	ACT	GAGGG	CAGG	ACGA	AGGGG	CTAG	TGGT	TCGCG	CAGG	ACCGG	115
QY	628	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	657
Db	1160	GG	ACAGCCC	AC	AGAACT	GTGA	AGCAT	TA	GAAG	AAAG	AGCAT	121
QY	658	GTG	GAGAC	CGT	TGGT	TCGCGG	ACGAG	CTCA	ATGGT	CAAGT	TTCACG	717
Db	1220	GTG	GAACCA	TGCT	TGGC	AGAC	CACTG	ATG	GCAG	AAATCC	ACGCA	127
QY	718	CAT	TATGCT	GAC	GTG	CTGG	CAAC	GGCG	CGCG	ACTCT	ACGGCAT	777
Db	1280	CAT	TACCTT	CTC	ACGT	TGTTT	TCGGT	GGC	AGCC	AGAT	TGTACA	133
QY	778	AAC	CCCAT	CAAC	ATCG	TTGTG	GTCA	AGGT	GTCTG	CTTCT	TAGAG	837
Db	1340	AAT	TAGTT	AG	CTT	GTG	TGGT	GA	AGAT	TTTGGT	CAATCC	139
QY	838	AAG	TCAC	CCG	CAAT	CGG	CCCTTGA	CGCT	TCGCA	ACTTCT	GTGCTT	897
Db	1400	GAA	GTAG	CCCTCA	ATGCT	GCCTT	CACTC	TGCG	GAATTT	TGCA	ACTT	145
QY	898	AAC	AAAGT	GAT	GAC	AAAG	CAAC	CCG	AGTACT	TGGG	ACAT	957
Db	1460	AAC	CCAC	CCAGT	AC	CGGGAT	G	CAGAC	CACTAT	GAC	ACAGCA	151
QY	958	GAC	CTGTG	CGAG	CCAC	CTGTG	AC	CCCTG	GGGAT	GGCTG	ATGTTGG	101
Db	1520	GAC	TTGTG	GGT	GGT	CCG	ACATG	TATCT	TTGG	ATGG	CTTGG	157
QY	1018	GAC	CCCAAG	AGA	AGCTG	CTGTG	CAAT	TAGG	ACGAT	GGGCTT	CCATCA	107
Db	1580	GAT	CCGAG	AGA	AGCTG	CTCC	GTAT	AGA	GATG	ATGGTT	TACAG	163
QY	1078	GCC	ACAG	AGT	GCG	CACTG	TTTCA	ACAT	ATGCC	CCATG	ACAAT	113
Db	1640	GCC	CATGAA	TAG	CCAC	CGTGT	TAA	CATGCC	ACATG	ATGCA	AAAGCA	169
QY	1138	GTG	TTTGG	GA	GTCC	GAG	CCAC	CA	CAATG	ATGTCCC	CAAC	119
Db	1700	CTT	AAATGG	TG	ACAG	GATTTCC	CA	ATG	TGG	CGCTCA	ATGTTT	175
QY	1198	GCC	AAAC	CCCTG	GTG	CTG	CACTG	CA	TCAT	CA	CCG	125
Db	1760	AGC	AGC	CTT	GGT	CTC	CTG	CA	CTG	CGCTT	ACAT	181
QY	1258	GGT	GACT	GT	CTCT	TGG	ACCA	AA	CCAG	CCCAT	CTCC	131

Db 1820 GGGGAATGTTGATGGACAAAGCCTCAGATCCCATACAGCTCCAGGCGCATCTCCCTGGC 1879

QY 1318 GCCAGCTACACCTGAGCCAGCAGTGGCGTGGCTTTTGGCTGGCTCCAGCCCTGT 1377

Db 1880 ACCTCGTACGATGCCAAGCGGAGTGGCAGTTTACATTTGGGAGGAGCTCCAAACACTGC 1939

QY 1378 CTTTACATGACGATAC---TGCACCAAGCTGTGTGTGACCGGGAAGGCCAAGGACAGATG 1434

Db 1940 CCGGATGACAGCCAGCAGATAGCAGCTTGTGGTGTACCGGCACCTCTGTGTGGGTGCTG 1999

QY 1435 GTGTGCCAGACCGCCGCTTCCCTGGCGATGGCCAGCTGTGGCGAGGCGCAAGCTC 1494

Db 2000 GTGTGTCAAACCAACACTTCCCGTGGGGATGGCCAGCTGTGGAGAAGGGAATGG 2059

QY 1495 TGCTCAAAAGGGCCCTGCTGGAGACACAACTCAACAAG-----CACAGGTGGAT 1548

Db 2060 TGTATCAACGGCAAGTGTGTGAACAAACCGACAGAAAGCATTTTGATACGGCTTTTCAT 2119

QY 1549 GGTTCCTGGCCCAATGGGATCCCTATGGCCCTGTGGCCACATGTGGTGGGGCGTG 1608

Db 2120 GGAAGCTGGGGAATGTGGGGCTTGGGGAGACTGTTCGAGAACGTGCGGTGGAGGATC 2179

QY 1609 CAGCTGGCCAGGAGCAGTCCACCAACCCACCCCTGCCAAGGGGGCAAGTACTGCGAG 1668

Db 2180 CAGTACAGATGAGGAATGTGACAAACCCAGTCCCAAGAAAGGAGGGAAGTACTGTGA 2239

QY 1669 GGAGTGAAGGTGAATACCGATCCTGCAACCTGGAGCCCTGCCCAAGCTCAGCCTCCGA 1728

Db 2240 GGCAACGAGTGGCTACAGATCCTGTAACTTGAGGACTGTCCAGAC---AATAATGA 2296

QY 1729 AAGAGCTTCGGGAGGAGCAGTGTGAGGCTTTCACGGCTACAAACAGCAGCACCACCG 1788

Db 2297 AAAACCTTTAGAGAGAAACAATGTGAAGCACACACAGAGTTTTCAAAAGCTTCCCTTGGG 2356

QY 1789 CTCACCTCGCGGTGGATGGTGGCCCAAGTACTCCGGGTGTCTCCCGGGACAAAGTGC 1848

Db 2357 AGTGGGCTCGGTGGAATGGATCCCAAGTACGCTGGCGTCCACCAAGGACAGGTGC 2416

QY 1849 AAGCTCATCTGCCAGACCAATGGCACTGGCTACTTCTATGTGTGGCACCCCAAGGTGTG 1908

Db 2417 AAGCTCATCTGCCAAGCAAGGCAATGGCTACTTCTGTTTGGCAGCCCAAGGTGTGA 2476

QY 1909 GACGGCAGCTGTGCTCTCTGACTCCAGCTCGTGTGTCTCAAGGCAAGTGCATCAAG 1968

Db 2477 GATGGTACTCCATGTAGCCAGATTCACCTCTGTGTGTGCAAGGACAGTGTGTAAA 2536

QY 1969 GCTGGCTGTGATGGGAACCTGGGCTCCAAGAAGAGATTCGACAAAGTGTGGGTGTGG 2028

Db 2537 GCTGGTGTGATCGCATCATAGACTCCAAAGAAAGTGTGTAATGTGGTGTGGCGG 2596

QY 2029 GGAGACAATAAGAGCTGCAAGAAGGTGACTGGACTCTTCCAAAGCCCATGATGGCTAC 2088

Db 2597 GGAATGGATCTACTTGTAAAAAATATCAGGATCAGTTACTAGTCAAAACCTGGATAT 2656

QY 2089 AATTTGTTGGTGGCCATCCCGGAGGCGCTCAAGCATCGACATCCGCCAGCGCGTTAC 2148

Db 2657 CATGATATCATCAATTCCAACTGGAGCCCAACATCGAAGTGAACAGCGGAACCCAG 2716

QY 2149 AAAGGGCTGATGGGATGACAACTACCTGGCTCTGAAGAACAGCCCAAGCAAGTACCTG 2208

Db 2717 AGGGATCCAGGAACAAATGGCAGCTTCTTGCCATCAAGCTGCTGATGGCATAATTT 2776

QY 2209 CTCAGGGGCAATTCGTGTGTGTCGGGGTGGAGCGGACCTGGTGTGAAGGCGAGTCTG 2268

Db 2777 CTTAATGGTGACTACACTTTGCTCCACTTAGAGCAAGACATTTATGTACAAAGGTGTGTC 2836

QY 2269 CTGGGTACAGCGGCGAGGCGCAGCGGTGGAGAGCCTCGAGGCTTCCCGGCCCATCCTG 2328

Db 2837 TTGAGGTACAGCGGCTTCTTGGCGCATTTGAAAGAAATTCGACGCTTTAGCCCTCTCAA 2896

QY 2329 GAGCCGCTGACCGTGGAGTCTCTCGGTGGGGAAGATGACACCGCCCGGTGCCGTAC 2388

Db 2897 GAGCCCTTGACCATCCAGGTCTTACTGTGGCAATGCCCTTCGACCTAAATTAATAC 2956

QY 2389 TCCTTCT 2395

Db 2957 ACCTACT 2963

Search completed: May 1, 2003, 22:35:26
Job time : 10450 secs

PA (KAZU-) KAZUSA DNA RES INST.


```
QY 1801 GTGCATGGGTGCCAAGTACTCCGGGTGTCTCCCGGACAAAGTCAAGCTCATCTGC 1860
Db 1801 GTGCATGGGTGCCAAGTACTCCGGGTGTCTCCCGGACAAAGTCAAGCTCATCTGC 1860
QY 1861 CGAGCCAATGGCACTGCTACTTCTATGTCTGCGACCCCAAGTGTGGAGCGCACGCTG 1920
Db 1861 CGAGCCAATGGCACTGCTACTTCTATGTCTGCGACCCCAAGTGTGGAGCGCACGCTG 1920
QY 1921 TGCTCTCTGACTCCACTCCGCTGTGTCCAAAGGAAGTGCATCAAGCTGGCTGTAT 1980
Db 1921 TGCTCTCTGACTCCACTCCGCTGTGTCCAAAGGAAGTGCATCAAGCTGGCTGTAT 1980
QY 1981 GGGAACTGGGCTCCAAAGAGATTCGCAAGTGTGGGTGTGGGGGAGACATAAG 2040
Db 1981 GGGAACTGGGCTCCAAAGAGATTCGCAAGTGTGGGTGTGGGGGAGACATAAG 2040
QY 2041 AGCTGCAAGAGTGTACTGACTCTTCCACCAAGCCCATGCATGGCTACAAATTCGTGGT 2100
Db 2041 AGCTGCAAGAGTGTACTGACTCTTCCACCAAGCCCATGCATGGCTACAAATTCGTGGT 2100
QY 2101 GCCATCCCGCAGCGCTCAAGCATCGACATCCGCGCGGCTTACAAAGGCTGATC 2160
Db 2101 GCCATCCCGCAGCGCTCAAGCATCGACATCCGCGCGGCTTACAAAGGCTGATC 2160
QY 2161 GGGATGACAACTACTGCTGTGAAGACAGCCAAAGGCAAGTACCTGCTCAACGGGCAT 2220
Db 2161 GGGATGACAACTACTGCTGTGAAGACAGCCAAAGGCAAGTACCTGCTCAACGGGCAT 2220
QY 2221 TTGCTGTGTGCGGTGGAGCGGACCTGTGTGTAAGGGCAGTCTGCTGGGTACAGC 2280
Db 2221 TTGCTGTGTGCGGTGGAGCGGACCTGTGTGTAAGGGCAGTCTGCTGGGTACAGC 2280
QY 2281 GGCACGGGCACAGCGTGGAGAGCCTGCAGGCTTCCCGGCCATCCTGAGCGCGTACAC 2340
Db 2281 GGCACGGGCACAGCGTGGAGAGCCTGCAGGCTTCCCGGCCATCCTGAGCGCGTACAC 2340
QY 2341 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGCTGCGTACTCTTCTATCTG 2400
Db 2341 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGCTGCGTACTCTTCTATCTG 2400
QY 2401 CCCAAAGAGCTCGGAGGACAAAGTCTCTCATCCCAAGAGCCCGCGGACCCCTCTGTC 2460
Db 2401 CCCAAAGAGCTCGGAGGACAAAGTCTCTCATCCCAAGAGCCCGCGGACCCCTCTGTC 2460
QY 2461 TTGCACAACAGCTCTCTCAGCTCTCCAAACAGGTGGAGCAGCGGACACAGGCGCCCT 2520
Db 2461 TTGCACAACAGCTCTCTCAGCTCTCCAAACAGGTGGAGCAGCGGACACAGGCGCCCT 2520
QY 2521 GCAGCTGGGTGCTGGCAGCTGGGGCGGTGCTCCGCGAGCTGCGGCAAGTGGCCTGCAG 2580
Db 2521 GCAGCTGGGTGCTGGCAGCTGGGGCGGTGCTCCGCGAGCTGCGGCAAGTGGCCTGCAG 2580
QY 2581 AAGCGGCGGTGACTCGCGGGCTCCGCGGGCAGCGACAGTCCCTGCTGTATGCA 2640
Db 2581 AAGCGGCGGTGACTCGCGGGCTCCGCGGGCAGCGACAGTCCCTGCTGTATGCA 2640
QY 2641 GCCCATCGGCGCTGGAGACACAAAGCTTCGGGGAGCCCTGCCCCACCTGGGAGCTCAGC 2700
Db 2641 GCCCATCGGCGCTGGAGACACAAAGCTTCGGGGAGCCCTGCCCCACCTGGGAGCTCAGC 2700
QY 2701 GCCTGTGCTACCTGTCTCAAGAGCTGCGGCGGGGATTTTCAGAGGCGTCACTCAAGTGT 2760
Db 2701 GCCTGTGCTACCTGTCTCAAGAGCTGCGGCGGGGATTTTCAGAGGCGTCACTCAAGTGT 2760
QY 2761 GTGGGCGCAGGAGCGGCTGTGGCGCGGACAGTGCACCTTGCACCGCAAGCCCCAG 2820
Db 2761 GTGGGCGCAGGAGCGGCTGTGGCGCGGACAGTGCACCTTGCACCGCAAGCCCCAG 2820
QY 2821 GAGCTGGACTTCTGCTCTCTGAGGCGGTGCTGA 2853
Db 2821 GAGCTGGACTTCTGCTCTCTGAGGCGGTGCTGA 2853
```

RESULT 2

AAD35569
ID AAD35569 standard; cDNA; 2853 BP.

XX AAD35569;

XX 26-JUL-2002 (first entry)

XX Human protease cDNA #2.

XX Human; novel human protein; NHP; protease; biological disorder; obesity;
high blood pressure; arthritis; connective tissue disorder; infertility;
gene therapy; enzyme; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..2853

FT /*tag= a

FT /product= "Human protease #2"

XX WO200226949-A2.

XX 04-APR-2002.

XX 27-SEP-2001; 2001WO-US30350.

XX 29-SEP-2000; 2000US-236689P.

XX (LEXI-) LEXICON GENETICS INC.

XX Friddle CJ, Hilbun E;

XX WPI; 2002-372123/40.

XX P-PSDB; AAD25541.

XX Novel nucleic acid encoding a human protease, useful as a hybridization probe for screening libraries and assessing gene expression patterns -

XX Claim 1; Page 35-36; 41pp; English.

XX The present sequence is a cDNA encoding novel human protein (NHP), human protease. NHPs share structural similarity with animal proteases particularly zinc metalloproteases. Sequences of the invention are useful in therapeutic, diagnostic and pharmacogenomic applications. NHP polynucleotides are used as hybridisation probes for screening libraries and assessing gene expression patterns. They can also be used for treating related biological disorders such as obesity, high blood pressure, arthritis, connective tissue disorders and infertility. They are also used in gene therapy.

XX SQ Sequence 2853 BP; 522 A; 920 C; 903 G; 508 T; 0 other;

Query Match 99.8%; Score 2846.6; DB 24; Length 2853;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2849; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGCTTTTGTCTGGGATCTTAACCTTTCGCGGGGCAACCGCTGGAGCTTTGAG 60

Db 1 ATGCTTCTGCTGGGATCTTAACCTTTCGCGGGGCAACCGCTGGAGCTTTGAG 60

QY 61 CCAGAGCGGAGGTAGTCTTCCATCCGACTGGACCGGACATTAACGGCTCCGCTAC 120

Db 61 CCAGAGCGGAGGTAGTCTTCCATCCGACTGGACCGGACATTAACGGCTCCGCTAC 120

QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTTCAGATCAGACATTT 180

Db 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTTCAGATCAGACATTT 180

QY 181 CAGGAGGACTTTTACCTACCTGACCGCGGATGCTCAGTTCTTGGCTCCGCGCTTCTCC 240

Db 181 CAGGAGGACTTTTACCTACCTGACCGCGGATGCTCAGTTCTTGGCTCCGCGCTTCTCC 240

Db 2401 CCCAAGAGGCTCGGAGGACAAAGTCTCTCATCCCAAGACCCCGGGAGCCTCTGTG 2460
Qy 2461 TTGCACACAGCGTCTCAGCTCTCCAAACAGGTGAGCAGCGGAGCAGAGGCCCT 2520
Db 2461 TTGCACACAGCGTCTCAGCTCTCCAAACAGGTGAGCAGCGGAGCAGAGGCCCT 2520
Qy 2521 GCACGCTGGGTGGCTGACAGCTGGGGCCGTGCTCCGCGAGCTGCGCAGTGGCTGCAG 2580
Db 2521 GCACGCTGGGTGGCTGACAGCTGGGGCCGTGCTCCGCGAGCTGCGCAGTGGCTGCAG 2580
Qy 2581 AAGCGCGGTGGAGTCCCGGGCTCCCGCGGAGCGCAGCGCTGCTGCTGATGCA 2640
Db 2581 AAGCGCGGTGGAGTCCCGGGCTCCCGCGGAGCGCAGCGCTGCTGCTGATGCA 2640
Qy 2641 GCCCATCGCCCTGGAGACACAGCTGCGGGGAGCCCTGCCCCACCTGGGAGCTCAGC 2700
Db 2641 GCCCATCGCCCTGGAGACACAGCTGCGGGGAGCCCTGCCCCACCTGGGAGCTCAGC 2700
Qy 2701 GCCTGGTACCTGCTCCAAAGAGTGGCGCCGGGATTTTCAGAGCGCTCACTCAAGTGT 2760
Db 2701 GCCTGGTACCTGCTCCAAAGAGTGGCGCCGGGATTTTCAGAGCGCTCACTCAAGTGT 2760
Qy 2761 GTGGGCCACGAGCGCGCTGCTGGCCCGGAGCCAGTGCACCTTGACCGCAAGCCGAG 2820
Db 2761 GTGGGCCACGAGCGCGCTGCTGGCCCGGAGCCAGTGCACCTTGACCGCAAGCCGAG 2820
Qy 2821 GAGTGGACTTCTGCTGCTGAGCGCTGCTGA 2853
Db 2821 GAGTGGACTTCTGCTGCTGAGCGCTGCTGA 2853

RESULT 3

AAD35571

ID AAD35571 standard; cDNA; 3446 BP.

AC AAD35571;

XX AAD35571;

DT 26-JUL-2002 (first entry)

DE Human protease cDNA #4.

KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme; ss.

XX Homo sapiens.

XX W0200226949-A2.

XX 04-APR-2002.

XX 27-SEP-2001; 2001WO-US30350.

XX 29-SEP-2000; 2000US-236689P.

XX (LEXI-) LEXICON GENETICS INC.

XX Friddle CJ, Hilbun E;

XX WPI; 2002-372123/40.

PT Novel nucleic acid encoding a human protease, useful as a hybridization
PT probe for screening libraries and assessing gene expression patterns
PS Disclosure; Page 40-41; 41pp; English.

CC The present sequence is a cDNA encoding novel human protein (NHP),
CC human protease. NHPs share structural similarity with animal proteases
CC particularly zinc metalloproteases. Sequences of the invention are
CC useful in therapeutic, diagnostic and pharmacogenomic applications.
CC NHP polynucleotides are used as hybridisation probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high

CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.

XX Sequence 3446 BP; 612 A; 1114 C; 1101 G; 619 T; 0 other;

Query Match 99.8%; Score 2846.6; DB 24; Length 3446;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2849; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATGCTTTTCTGGGATCTTAACCTTTCGCGCGGCGAACCCTTGGAGGCTTTTTCAG 60

Db 397 ATGCTTCTCTGGGATCTTAACCTTTCGCGCGGCGAACCCTTGGAGGCTTCTGAG 456

Qy 61 CCAGAGCGGAGTCTAGTCTTCCATCCGACTGACCGCGGACATTAACGCGCGCTTAC 120

Db 457 CCAGAGCGGAGTCTAGTCTTCCATCCGACTGACCGCGGACATTAACGCGCGCTTAC 516

Qy 121 TACTGGCGGGTCCCGAGGACTCGGGGATCAGGAGCTCATTTTTCAGATCAACAATTT 180

Db 517 TACTGGCGGGTCCCGAGGACTCGGGGATCAGGAGCTCATTTTTCAGATCAACAATTT 576

Qy 181 CAGGAGGACTTTTACCTACCTGACCGCGGATGCTAGTCTTGGCTCCGCTTCTCC 240

Db 577 CAGGAGGACTTTTACCTACCTGACCGCGGATGCTAGTCTTGGCTCCGCTTCTCC 636

Qy 241 ACTGAGCATCTGGGCTCCCTCCAGGGGCTTCACGGGGGCTTTCAGACCTGCGAGCG 300

Db 637 ACTGAGCATCTGGGCTCCCTCCAGGGGCTTCACGGGGGCTTTCAGACCTGCGAGCG 696

Qy 301 TCGTTCTATTCTGGGAGCTGAACCGCGAGCGGACTCGTTGGCTGCTGAGCTGTGTC 360

Db 697 TCGTTCTATTCTGGGAGCTGAACCGCGAGCGGACTCGTTGGCTGCTGAGCTGTGTC 756

Qy 361 GGGGGCTCCGCGAGCTTGGCTACCGAGCGCGGAGTATGTCATTAACCGCTGCCC 420

Db 757 GGGGGCTCCGCGAGCTTGGCTACCGAGCGCGGAGTATGTCATTAACCGCTGCCC 816

Qy 421 AATGCTAGCGCGCGCGGCGCAGCAACAGCGGCGCACACTTCTCCAGCGCGG 480

Db 817 AATGCTAGCGCGCGCGGCGCAGCAACAGCGGCGCACACTTCTCCAGCGCGG 876

Qy 481 GGTGTTCCGGCGCGCTTCGCGAGACCCCACTCTCGCTCGGGGTGGCTCGGGCTGG 540

Db 877 GGTGTTCCGGCGCGCTTCGCGAGACCCCACTCTCGCTCGGGGTGGCTCGGGCTGG 936

Qy 541 AACCCCGCATCTACGGGCTTGACACCTTACAAAGCGCGCGCGCGGCTTCGGGAG 600

Db 937 AACCCCGCATCTACGGGCTTGACACCTTACAAAGCGCGCGCGCGGCTTCGGGAG 996

Qy 601 AGTCGTAGCGCGCGAGTCTGGGCGCGCAAGCGCTTTCGTCTATCCCGGCTACGTCG 660

Db 997 AGTCGTAGCGCGCGAGTCTGGGCGCGCAAGCGCTTTCGTCTATCCCGGCTACGTCG -1056

Qy 661 GAGACGCTGGGTGCGGAGCAGTCAATGTTCAGTTCCACGGCGCGGACCTGGAACAT 720

Db 1057 GAGACGCTGGGTGCGGAGCAGTCAATGTTCAGTTCCACGGCGCGGACCTGGAACAT 1116

Qy 721 TATCTGCTGACGCTGCTGCAACGCGCGGAGCTTACCGCATCCCGAGCATCCTCAAC 780

Db 1117 TATCTGCTGACGCTGCTGCAACGCGCGGAGCTTACCGCATCCCGAGCATCCTCAAC 1176

Qy 781 CCATCAACATCGTTGTTGTTCAAGGTGCTGCTTCTTAGAGATCGTACTCCTGGGCCCAAG 840

Db 1177 CCATCAACATCGTTGTTGTTCAAGGTGCTGCTTCTTAGAGATCGTACTCCTGGGCCCAAG 1236

Qy 841 GTCACCGCATCGGGCCCTGACGCTGCGCAACTTCTGTGCTGCGCAGAAAGCTGAAC 900

Db 1237 GTCACCGCATCGGGCCCTGACGCTGCGCAACTTCTGTGCTGCGCAGAAAGCTGAAC 1296

Qy 901 AAAGTGTGACAAACACCCCGAGTCTGGGAGCTGCATCCCTCTTACCTAGCAGAG 960

Db 1297 AAAGTGTGACAAACACCCCGAGTCTGGGAGCTGCATCCCTCTTACCTAGCAGAG 1356

QY 961 CTGTGTGGAGCCACACCTGTGTACACCCCTGGGCGATGGCTGATGTGGGTACCATGTGTGAC 1020
Db 1357 CTGTGTGGAGCCACACCTGTGTACACCCCTGGGCGATGGCTGATGTGGGTACCATGTGTGAC 1416
QY 1021 CCCAAGAGAAGCTGCTGTCTGTCATTGAGGAGGATGGCTTCCCATCAGCCCTTACCACTGCC 1080
Db 1417 CCCAAGAGAAGCTGCTGTCTGTCATTGAGGAGGATGGCTTCCCATCAGCCCTTACCACTGCC 1476
QY 1081 CAGGAGTGGGCGACAGTGTTCACATGCCCCATGACAACTGTGAAAGTCTGTGAGGAGGTG 1140
Db 1477 CAGGAGTGGGCGACAGTGTTCACATGCCCCATGACAACTGTGAAAGTCTGTGAGGAGGTG 1536
QY 1141 TTTGGGAGCTCGAGCCAAACACATGATGTGTCCCCGACCCCTCATCCAGATCGACCTGCC 1200
Db 1537 TTTGGGAGCTCGAGCCAAACACATGATGTGTCCCCGACCCCTCATCCAGATCGACCTGCC 1596
QY 1201 AACCCCTGTGTCAGCTGCTGCATCATCACCGACTTCTTGACACGCGGCACGGT 1260
Db 1597 AACCCCTGTGTCAGCTGCTGCATCATCACCGACTTCTTGACACGCGGCACGGT 1656
QY 1261 GACTGCTCTGTGACCAACCCAGCAAGCCCATCTCCCTGCCCGAGGATCTGCCGGCGCC 1320
Db 1657 GACTGCTCTGTGACCAACCCAGCAAGCCCATCTCCCTGCCCGAGGATCTGCCGGCGCC 1716
QY 1321 AGCTACACCTGAGCAGCAGTCCGAGCTGGCTTTTGGGCTGGGCTCAAGCCCTGTCT 1380
Db 1717 AGCTACACCTGAGCAGCAGTCCGAGCTGGCTTTTGGGCTGGGCTCAAGCCCTGTCT 1776
QY 1381 TACATGCACTACTGCAACCAAGCTGTGTGCACCGGAAGGCCAAGGACAGATGTTGTGC 1440
Db 1777 TACATGCACTACTGCAACCAAGCTGTGTGCACCGGAAGGCCAAGGACAGATGTTGTGC 1836
QY 1441 CAGACCCGCACTTCCCTTGGGCGGATGAGCAAGCTGTGGGAGGGCAAGCTCTGCCCTC 1500
Db 1837 CAGACCCGCACTTCCCTTGGGCGGATGAGCAAGCTGTGGGAGGGCAAGCTCTGCCCTC 1896
QY 1501 AAAGGGCCCTGCTGGAGACACACCTCAACACACAGAGTGTGATGTTCTGGGCC 1560
Db 1897 AAAGGGCCCTGCTGGAGACACACCTCAACACACAGAGTGTGATGTTCTGGGCC 1956
QY 1561 AAATGGGATCCCTATGCCCCCTGCTCGCGCACATGTGTGGGGCGTGCAGCTGGCCAGG 1620
Db 1957 AAATGGGATCCCTATGCCCCCTGCTCGCGCACATGTGTGGGGCGTGCAGCTGGCCAGG 2016
QY 1621 AGGAGTGCACCAACCCACCCCTGCCAAGGGGGGCAAGTACTGCAGGAGTGGAGGTG 1680
Db 2017 AGGAGTGCACCAACCCACCCCTGCCAAGGGGGGCAAGTACTGCAGGAGTGGAGGTG 2076
QY 1681 AAATACCGATCTTGCACCTTGGAGCCCTGCCCCAGCTCAGCCCTCGGGAAGAGCTTCGGG 1740
Db 2077 AAATACCGATCTTGCACCTTGGAGCCCTGCCCCAGCTCAGCCCTCGGGAAGAGCTTCGGG 2136
QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGACACCAACCCGGCTCACTCTCGCC 1800
Db 2137 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGACACCAACCCGGCTCACTCTCGCC 2196
QY 1801 GTGCGATGGTGCCCAAGTACTCCGGGCTGTCTCCCGGGGACAAGTGCAGAGCTCATCTGC 1860
Db 2197 GTGCGATGGTGCCCAAGTACTCCGGGCTGTCTCCCGGGGACAAGTGCAGAGCTCATCTGC 2256
QY 1861 CGAGCCAAATGGCACTGGCTACTTCTATGTCTGGCACCCCAAGTGTGGACGCGAGCTG 1920
Db 2257 CGAGCCAAATGGCACTGGCTACTTCTATGTCTGGCACCCCAAGTGTGGACGCGAGCTG 2316
QY 1921 TGCTCTCTGACTCCACCTCCGCTGTGTGTCCAAAGGCAAGTGCATCAAGGCTGGCTGTGAT 1980
Db 2317 TGCTCTCTGACTCCACCTCCGCTGTGTGTCCAAAGGCAAGTGCATCAAGGCTGGCTGTGAT 2376
QY 1981 GGGAACTGGGCTCCAAAGAGATTCGACAACTGTGGGTGTGTGGGGAGACAATAAG 2040
Db 2377 GGGAACTGGGCTCCAAAGAGATTCGACAACTGTGGGTGTGTGGGGAGACAATAAG 2436
QY 2041 AGCTGCAAGAAAGTGTGACTGGACTTTCACCAACGCCCATGCATGGCTTACAATTCGTGGTG 2100

Db 2437 AGCTSCAAGAAGGTGACTGAGACTTCTTCCAAAGCCCATGCAATGCTGCTGCTGGTG 2496
QY 2101 GCATCTCCCGCAGGCGCTCAAGCATCGACATCCGCAGCGCGTTCACAAAGGCTGATC 2160
Db 2497 GCATCTCCCGCAGGCGCTCAAGCATCGACATCCGCAGCGCGTTCACAAAGGCTGATC 2556
QY 2161 GGGGATGACAACTACCTGGCTCTGAAGAACAGCAAGTACCTGCTCAACGGGCAT 2220
Db 2557 GGGGATGACAACTACCTGGCTCTGAAGAACAGCAAGTACCTGCTCAACGGGCAT 2616
QY 2221 TTCGTGTGTCTGGCGGTGGAGCGGACCTGTGTGAGGGCAGCTCTGCTCGGCTACAGC 2280
Db 2617 TTCGTGTGTCTGGCGGTGGAGCGGACCTGTGTGAGGGCAGCTCTGCTCGGCTACAGC 2676
QY 2281 GSCAGGGCAGACGCGTGGAGAGCTGTCAGGCTTCCCGGCCCATCTCTGGAGCCCTGACC 2340
Db 2677 GSCAGGGCAGACGCGTGGAGAGCTGTCAGGCTTCCCGGCCCATCTCTGGAGCCCTGACC 2736
QY 2341 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCCGCTACTCTTCTATCTG 2400
Db 2737 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCCGCTACTCTTCTATCTG 2796
QY 2401 CCCAAGAGCCCTCGGAGGAGCAAGTCTCTCAACAGAGTGGAGCAGCCGAGCAGCCCT 2520
Db 2857 TTGCACAACAGCGTCTCAGCCTCTCCAAACAGAGTGGAGCAGCCGAGCAGCCCT 2916
QY 2521 GCACCTGGGTGGCTGGAGCTGGGGCGCTGCTCCCGGAGCTGGGAGTGGGCTGCAG 2580
Db 2917 GCACCTGGGTGGCTGGAGCTGGGGCGCTGCTCCCGGAGCTGGGAGTGGGCTGCAG 2976
QY 2581 AAGCGGGCGGTGGACTGCGGGGCTCCCGCGGAGCAGCGCTGCTGCTGTGATGCA 2640
Db 2977 AAGCGGGCGGTGGACTGCTGGGGCTCCCGCGGAGCAGCGCTGCTGCTGTGATGCA 3036
QY 2641 GCCATCGCCCGGTGGAGACACAGCTTGGGGGAGCCCTGCCCCACCTGGGAGCTCAGC 2700
Db 3037 GCCATCGCCCGGTGGAGACACAGCTTGGGGGAGCCCTGCCCCACCTGGGAGCTCAGC 3096
QY 2701 GCTCGTCACTCTGCTCCAAAGAGCTGCGCGCGGGATTTTCAGAGCGCTCACTCAAGTGT 2760
Db 3097 GCTCGTCACTCTGCTCCAAAGAGCTGCGCGCGGGATTTTCAGAGCGCTCACTCAAGTGT 3156
QY 2761 GTGGGCCACGGAGCGGCTGCTGGCCCGGAGCAGTGCACCTTGCACCGCAAGCCCAG 2820
Db 3157 GTGGGCCACGGAGCGGCTGCTGGCCCGGAGCAGTGCACCTTGCACCGCAAGCCCAG 3216
QY 2821 GAGCTGGAGTCTTGTGGTCTCTGAGCCGCTGCTGA 2853
Db 3217 GAGCTGGAGTCTTGTGGTCTCTGAGCCGCTGCTGA 3249
RESULT 4
AAS97182
ID AAS97182 standard; cDNA; 2937 BP.
XX
AC AAS97182;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human metalloprotease partial DNA sequence #11.
KW Human; protease; PCR primer; cytostatic; immunomodulator, cardiant;
KW vasotropic; antinigraine; analgesic; endocrine; nootropic; tranquiliser;
KW hyperensive; hypotensive; neuroleptic; neuroprotective; anabolic;
KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;
KW metalloprotease; serine protease; cancer; haematopoietic; breast;
KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
KW immune-related disease; cardiovascular disease; neuronal disease;

KW migraine; sexual dysfunction; mood disorder; attention disorder;
 KW cognition disorder; hypotension; hypertension; psychotic disorder;
 KW dyskinesia; metabolic disorder; inflammatory disorder; ss.

OS Homo sapiens.

PN WO200183782-A2.

XX 08-NOV-2001.

XX 04-MAY-2001; 2001WO-US14431.

XX 04-MAY-2000; 2000US-201879P.

XX (SUGE-) SUGEN INC.

XX Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 PI Payne V;

XX WPI; 2002-041502/05.
 DR P-PSDB; AAU72899.

PT Novel protease polypeptide useful for screening for substances that may
 be used to treat, e.g., cancers, immune-related diseases,
 PT cardiovascular disease, migraine, pain, psychotic and inflammatory
 PT disorders -

XX Claim 30; Figure 1P; 232pp; English.

XX The invention relates to an isolated, enriched, or purified protease
 CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
 CC screen for substances (S) that may modulate its activity. Administering
 CC S (which modulates protease activity in vitro) may be used to treat a
 CC disease or disorder selected from cancers (e.g., of tissues, of blood or
 CC hematopoietic origin, of the breast, colon, lung, prostate, cervical,
 CC brain, ovarian, bladder or kidney), immune-related diseases and
 CC disorders, cardiovascular disease, brain or neuronal-associated diseases
 CC (e.g., central or peripheral nervous system diseases, migraine, pain,
 CC sexual dysfunction, mood disorders, attention disorders, cognition
 CC disorders, hypotension, hypertension, psychotic disorders, neurological
 CC disorders and dyskinesias), metabolic disorders and inflammatory
 CC disorders. (I) may also be useful as a diagnostic tool for a disease or
 CC disorder such as those above. AAS97159-AAS97195 represent human
 CC protease coding sequences and primers of the invention.

XX Sequence 2937 BP; 531 A; 958 C; 936 G; 512 T; 0 other;

Query Match 98.6%; Score 2814.4; DB 24; Length 2937;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 2847; Conservative 0; Mismatches 6; Indels 9; Gaps 2;

QY 1 ATGCTTTTGTGGGCAATCCCTGAGCTTTCGCGGGCGAACCCTGGAGGCTTTGAG 60

Db 76 ATGCTTCTGTGGGCAATCCCTGAGCTTTCGCGGGCGAACCCTGGAGGCTTTGAG 135

QY 61 CCAGAGCGGAGTAGTCTGCTCCATCCGACTGGACCGGACATTAACGGCGCGGCTAC 120

Db 136 CCAGAGCGGAGTAGTCTGCTCCATCCGACTGGACCGGACATTAACGGCGCGGCTAC 195

QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTCAGATCACAGATT 180

Db 196 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTCAGATCACAGATT 255

QY 181 CAGGAGACTTTTACCTACACTGACCGGATGCTCAGTCTTGCTCCCGCTTCTCC 240

Db 256 CAGGAGACTTTTACCTACACTGACCGGATGCTCAGTCTTGCTCCCGCTTCTCC 315

QY 241 ACTGAGACTTGGGCTCCCGCTCCAGGGCTCACCGGGGCTCTTCAGACTGGAGCG 300

Db 316 ACTGAGACTTGGGCTCCCGCTCCAGGGCTCACCGGGGCTCTTCAGACTGGAGCG 375

QY 301 TCGTCTTATCTGGGAGCTGAACCGGAGCGGAGCTCTCGCTGTGAGCTGTGC 360

Db 1456 TACATGCACTACTGCACCAAGCTGTGGTGCACCGGGAAGGCAAGGACAGATGGTGTGC 1515

Db 376 TGCTTCTATTCTGGGACGTGAACCGGAGCGGAGCTCGTTGCTGCTGTGAGCCTGTGC 435

QY 361 GGGGGGCTCCGGGAGCCTTTGGCTACCGAGCGCGGAGTATGTATAGCCCGTGTGCC 420

Db 436 GGGGGGCTCCGGGAGCCTTTGGCTACCGAGCGCGGAGTATGTATAGCCCGTGTGCC 495

QY 421 AATGCTAGCGCGCGGCGCAGCGCAACAGCGGCGGAGCTCTTCTCCAGTCCGCGG 480

Db 496 AATGCTAGCGCGCGGCGCAGCGCAACAGCGGCGGAGCTCTTCTCCAGGCGCGG 555

QY 481 GGTGTTCCGGGCGGCGCTTCCGGAGACCCACTCTCGCTGCGGGGTGCTTCGCGCTGG 540

Db 556 GGTGTTCCGGGCGGCGCTTCCGGAGACCCACTCTCGCTGCGGGGTGCTTCGCGCTGG 615

QY 541 AACCCCGCATCTTACGGGCGCTTGAACCGCGCGGCGGCGGCGGCTTCGCGGAG 600

Db 616 AACCCCGCATCTTACGGGCGCTTGAACCGCGCGGCGGCGGCGGCTTCGCGGAG 675

QY 601 AGTCGTAGCGCGCGGCGCTTCCGGCGGCGGCGGCGGCGGCTTCTATCCCGGGTACGTG 660

Db 676 AGTCGTAGCGCGCGGCGCTTCCGGCGGCGGCGGCGGCGGCTTCTATCCCGGGTACGTG 735

QY 661 GAGAGCTGGTGGTCCGGAGCAGTCAATGTCAAGTTCACGGCGCGGAGCTTGAACAT 720

Db 736 GAGAGCTGGTGGTCCGGAGCAGTCAATGTCAAGTTCACGGCGCGGAGCTTGAACAT 795

QY 721 TATCTGTGACGCTGCTGGCAACGGCGGCGGCGGCGGCGGCTTACCGCCATCCAGCATCTCAAC 780

Db 796 TATCTGTGACGCTGCTGGCAACGGCGGCGGCGGCGGCTTACCGCCATCCAGCATCTCAAC 855

QY 781 CCATCAACATGTTGTTGTTCAAGTGTGCTTCTTAGAGATCGTGACTCCGGGCGGCGG 840

Db 856 CCATCAACATGTTGTTGTTCAAGTGTGCTTCTTAGAGATCGTGACTCCGGGCGGCGG 915

QY 841 GTCACGGCAATGCGGCGCTGAGCGTGGCGCAACTTCTGTGCTGGCAGAGAAGTGAAC 900

Db 916 GTCACGGCAATGCGGCGCTGAGCGTGGCGCAACTTCTGTGCTGGCAGAGAAGTGAAC 975

QY 901 AAAGTGAAGTGAACAGCACCCGAGTGTGGAGACTGCATCTTCCAGCAGGAGGAG 960

Db 976 AAAGTGAAGTGAACAGCACCCGAGTGTGGAGACTGCATCTTCCAGCAGGAGGAG 1035

QY 961 CTGTGTGGAGCACCACTGTGACACCTTGGGCGATGCTGTGGTGTGAGTGTGAGTGTGAC 1020

Db 1036 CTGTGTGGAGCACCACTGTGACACCTTGGGCGATGCTGTGGTGTGAGTGTGAGTGTGAC 1095

QY 1021 CCCAGAGAGCTGCTGTGTCATTGAGAGCATGGGCTTCCATAGACCTTCCAGCATGCC 1080

Db 1096 CCCAGAGAGCTGCTGTGTCATTGAGAGCATGGGCTTCCATAGACCTTCCAGCATGCC 1155

QY 1081 CACGAGCTGGGCGGCTGTTCAACATGCCCATGCAATGTGAAGTCTGTGAGGAGTG 1140

Db 1156 CACGAGCTGGGCGGCTGTTCAACATGCCCATGCAATGTGAAGTCTGTGAGGAGTG 1215

QY 1141 TTTTGGAAAGTCCGAGCAACACATGATGTCCCGACCCCTCATCCAGATCGACCTGTGCC 1200

Db 1216 TTTTGGAAAGTCCGAGCAACACATGATGTCCCGACCCCTCATCCAGATCGACCTGTGCC 1275

QY 1201 AACCCCTGGTCAGCTGAGTGTGCCATCATCCGAGTCTTCTTGGAGAGTGGGACGGT 1260

Db 1276 AACCCCTGGTCAGCTGAGTGTGCCATCATCCGAGTCTTCTTGGAGAGTGGGACGGT 1335

QY 1261 GACTGCTCTCTGGAGCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320

Db 1336 GACTGCTCTCTGGAGCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1395

QY 1321 AGCTACACCTTGAGCGAGCAGTGGAGTGGCTTTTGGCGTGGGCTTCCAGCCCTGTCT 1380

Db 1396 AGCTACACCTTGAGCGAGCAGTGGAGTGGCTTTTGGCGTGGGCTTCCAGCCCTGTCT 1455

QY 1381 TACATGCACTACTGCACCAAGCTGTGGTGCACCGGGAAGGCAAGGACAGATGGTGTGC 1440

Db 1456 TACATGCACTACTGCACCAAGCTGTGGTGCACCGGGAAGGCAAGGACAGATGGTGTGC 1515

Qy	2512	AGGCCCCCTGCACGCTGGGTGGCTGGCAGCTGGGGGGCGTGCTCCCGAGCTGCCGGCAGT	2571		
Db	2596	AGGCCCCCTGCACGCTGGGTGGCTGGCAGCTGGGGGGCGTGCTCCCGAGCTGCCGGCAGT	2655		
Qy	2572	GGCCTCAGAAAGCGGGGTGGACTGCCGGGGGCTCCGCCGGGCAGCGCACGGTCCCTGCC	2631		
Db	2656	GGCCTCAGAAAGCGGGGTGGACTGCCGGGGGCTCCGCCGGGCAGCGCACGGTCCCTGCC	2715		
Qy	2632	TGTGTATCAGCCCATCGGCCCTGTGGAGACACAAAGCTTGCGGGGAGCGCTTGCACCTTGG	2691		
Db	2716	TGTGTATCAGCCCATCGGCCCTGTGGAGACACAAAGCTTGCGGGGAGCGCTTGCACCTTGG	2775		
Qy	2692	GAGCTCAGCGCTGTGACCTTGTCTCCAAAGAGCTCGGGCCGGGGATTTCAGAGCGCTCA	2751		
Db	2776	GAGCTCAGCGCTGTGACCTTGTCTCCAAAGAGCTCGGGCCGGGGATTTCAGAGCGCTCA	2835		
Qy	2752	CTCAAGTGTGTGGGCGCAGGAGCGCGGTGCTGTGGCCCGGAGCAGCTGCACCTTGCACCGC	2811		
Db	2836	CTCAAGTGTGTGGGCGCAGGAGCGCGGTGCTGTGGCCCGGAGCAGCTGCACCTTGCACCGC	2895		
Qy	2812	AAGCCCCAGAGCTGGACTTCTGGTCTCTGAGGCCGTGCTGA	2853		
Db	2896	AAGCCCCAGAGCTGGACTTCTGGTCTCTGAGGCCGTGCTGA	2937		
RESULT 5					
ABK12894					
ID	ABK12894 standard; cDNA; 2930 BP.				
AC	ABK12894;				
XX					
DT	09-APR-2002 (first entry)				
XX					
DE	Human protease PRTS-11 cDNA sequence.				
XX					
KW	Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;				
KW	cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;				
KW	inflammatory disorder; acquired immunodeficiency syndrome; AIDS;				
KW	cell proliferative disorder; developmental disorder; epilepsy;				
KW	Duchenne muscular dystrophy; epithelial disorder; neurological disorder				
KW	reproductive disorder; endometriosis; ss.				
XX					
OS	Homo sapiens.				
XX					
FH	Key	Location/Qualifiers			
FT	CDS	75..2930			
FT		/*tag= a			
FT		/partial			
FT		/product= "Human protease PRTS-11"			
FT		/note= "This sequence lacks a stop codon"			
XX					
XX	WO2001198468-A2.				
XX					
PD	27-DEC-2001.				
XX					
PF	13-JUN-2001;	2001WO-US19178.			
XX					
PR	16-JUN-2000;	2000US-212336P.			
PR	22-JUN-2000;	2000US-213955P.			
PR	29-JUN-2000;	2000US-215396P.			
PR	07-JUL-2000;	2000US-216821P.			
PR	14-JUL-2000;	2000US-218946P.			
XX					
PA	(INCY-) INCYTE GENOMICS INC.				
XX					
PI	Yue H, Elliott VS, Gandhi AR, Lal P, Au-Young J, Tribouley CM;				
PI	Delegane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;				
PI	Walila NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;				
PI	Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;				
PI	Kallick DA;				
XX					
DR	WPI; 2002-090437/12.				
DR	P-PSDB; AAU74751.				

Db 97 GCGAGCCAGACTCCTTTGCTGCTTAAGCCTATGCGGGGTCTCCGCGAGCCTTTGGC 156
Qy 385 TACGAGCGCGAGTAGTATAGCCGCTGCCCCAATGCTAGCGCGCGCGGCGCAG 444
Db 157 TACCAAGGTGCGAGTAGTATAGCCCTTGTGCCAACACACGCGCGCTGAGGCGCAG 216
Qy 445 CGCAACAGCCAGGCGCACACCTTCTCCAGCGCGGGGTGTTCGGCGCGCCCTTCGGGA 504
Db 217 CGTCATAGCCAGGCGCACACCTTCTCCAGCGCGGGGTGCTCCCGTAGGCGCTTCGGGA 276
Qy 505 GACCCCACTCTCGCTCGGGGTGGCTCGGGTGGAAACCCCGCATCTTACGGGCGCTG 564
Db 277 GACCCCTACCTCTCGCTCGGGGTGGCTCGGGTGGAAACCCCGCATCTTACGGGCGCTG 336
Qy 565 GACCCCTTACAAGCGCGCGGGGTGCTCGGGGTGGAGTCTGAGCGCGCAGCTCTGGG 624
Db 337 GACCCCTTATAACACCGCGGAGCGGGGTGCGGGGTGGAAAGCCACACCGCGCGAGCTCTGGG 336
Qy 625 CGCGCAAGCGTTTCTGCTATCCCGCGGTACGTGGAGACGCTGGTGGTGGCGGCGAG 684
Db 397 CGCGCAAGCGCTTCTGCTATACACCGGTACGTGGAGACACTGGTGGTGGCGGAGAG 456
Qy 685 TCAATGCTCAAGTTTCAACGCGCGGACCTGGAACATATCTGCTAGCGTGTGGCAACG 744
Db 457 TCAATGCTCAAGTTTCAACGCGCGGATTTGGAACATATCTGCTAGCGTGTGGCAACG 516
Qy 745 GCGCGCGACTCTACCGCCATCCAGCATCTCAACCCCAATCAACATCTGTTGGTCAAG 804
Db 517 GCGCGCGACTCTACCGCCACCCAGCATCTCAACCCATCAACATCTGTTGGTCAAG 576
Qy 805 GTGCTGCTTCTAGAGTCTGACTCGGGGCCCAAGTCAACCGCAATCGGGCCCTGAGC 864
Db 577 GTGTTACTCTTAGAGATCTGACATGCGGCCCAAGTCAACCGCAATCGGGCCCTGACT 636
Qy 865 CTGCGCAACTCTGCTGCGGAGAGAGCTCAACAAAGTGAAGTCAACAGCACCCCGAG 924
Db 637 CTGCGCAACTCTGCTGCGGAGAGAGTGAACAAAGTGAACAGCACCCCGAG 696
Qy 925 TACTGGACACTGCGATCTTCAACAGCGAGGACTGTGTGGAGCCACCACTGTGAC 984
Db 697 TACTGGACACTGCGATCTTCAACAGCGAGGACTGTGTGGAGCCACCACTGTGAC 756
Qy 985 ACCTGGGATGCTGATGTGGTACCATCTGTGACCCCAAGAGAGCTGCTGTGCTAT 1044
Db 757 ACCTGGGATGCTGATGTGGGACCATGTGTGATCCCAAGAGAGCTGCTGTGCTATC 816
Qy 1045 GAGGACATGGGCTTCCATCAGCTTCAACACTGCGCCAGCGCTGGGCGCACCTGTTCAC 1104
Db 817 GAGGACATGGGCTTCCGTCGGCTTCAACACTGCGCCATGAGCTGGGCGCATGTGTCAC 876
Qy 1105 ATGCCCATGACAATGTGAAGTCTGTGAGGAGGTGTTTGGAGCTCCGACCAACAC 1164
Db 877 ATGCCCATGACAACGTGAAGTGTGTGAGGAGGTGTTTGGAGCTCAGACCAACAC 936
Qy 1165 ATGATGTCCCGACCTCTCAGATCGACCTGCGCAACCCCTGTGACCTTGCAGTGT 1224
Db 937 ATGATGTCTCCGACACTCATCCAGATCGACCTGCGCAACCCCTGTGACCTTGCAGTGT 996
Qy 1225 GCATCATCACCAGCTTCTTGGACAGCGGCGACGCTGCTGCTGCTGACCAACCCAGC 1284
Db 997 GCATCATCACCAGCTTCTTGGACAGCGGCGACGCTGCTGCTGCTGACCAACCCAGC 1056
Qy 1285 AAGCCATCTCTCCGCGAGATCTGCGGGCGGCGAGTACACCTGAGCGAGTGC 1344
Db 1057 AAGCCATCTCTCCGCGAGTCTGAGGACCTTGCAGGCGACACATGTTGAGCAACAGTGC 1116
Qy 1345 GAGCTGGCTTTGGGCTGGGCTCCAGCCCTGCTTACATCAGTACTGACCAAGCTG 1404
Db 1117 GAGCTGGCTTTGGGCTGGGCTTAAAGCCCTGCCATATATGACTGACTGTACAAGCTG 1176
Qy 1405 TGSTGACCGGGAAGGCCAAGGACAGATGTTGTGCGACACCCGACCTTCCCTGGGCC 1464
Db 1177 TGSTGACCGGGAAGGCCAAGGCGAGATGTTGTGCGACACTCGCCACTTCCCTGGGCA 1236

Qy 1465 GATGGCACAGCTGTGGCGAGGCAAGCTCTGCCTCAAGGGGCTGCGTGAGAGACAC 1524
Db 1237 GATGGCACAGCTGTGGTGAGGCAAGTCTGCCTCAAGGGAGCTGCGTGAGAGACAC 1296
Qy 1525 AACCTCAACAAGCAGAGGTGATGTTCTCTGGGCAATGGGATCCCTATGGCTCTGTC 1584
Db 1297 AACCCAAACAGTACCGGTGGACGCCCTTGGGCCAAGTGGGAGCCCTATATGGTCCCTGC 1356
Qy 1585 TCAGCACAATGTGGTGGGCGCTGACGCTGGCCAGGAGGCA--GTGCACCAACCCACCC 1642
Db 1357 TCAGCACAATGTGGTGGGCGCGCAGCTGGCCCGGAGGAGTGCAGCAACCTACCC 1416
Qy 1643 CTGCCAACGGG--GGCAAGTACTGCGAGGAGTGAAGTGAATACCGATCTCTGCAACCTG 1701
Db 1417 CTGCCAACGGGCGGGAAGTACTGCGAGGAGTGAAGTGAATACCGATCTCTGCAACCTG 1476
Qy 1702 GAGCCCTGCGCCAGCTCAGCCCTCGGAAAGAGCTTCCCGGA 1742
Db 1477 GAGCCCTGCGCCAGCTCAGCCCTGCGAAGAGCTTCCCGGA 1517

RESULT 7
AAD35570

ID AAD35570 standard; cDNA; 1104 BP.

XX AAD35570;

XX 26-JUL-2002 (first entry)

XX Human protease cDNA #3.

XX Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
PH 1.1104
FT CDS
FT /*tag= a
FT /product= "Human protease #3"

XX WO200226949-A2.

XX 04-APR-2002.

XX 27-SEP-2001; 2001WO-US30350.

XX 29-SEP-2000; 2000US-236689P.

XX (LEXI-) LEXICON GENETICS INC.

XX Friddle CJ, Hilbun E;

XX WPI; 2002-372123/40.

XX P-PSDB; AAE22542.

XX Novel nucleic acid encoding a human protease, useful as a hybridization
probe for screening libraries and assessing gene expression patterns -
Disclosure; Page 38-39; 41pp; English.XX The present sequence is a cDNA encoding novel human protein (NHP),
human protease. NHPs share structural similarity with animal proteases
particularly zinc metalloproteases. Sequences of the invention are
useful in therapeutic, diagnostic and pharmacogenomic applications.
XX NHP polynucleotides are used as hybridisation probes for screening
libraries and assessing gene expression patterns. They can also be
used for treating related biological disorders such as obesity, high
blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy.

XX

Query Match 36.6%; Score 1045.2; DB 21; Length 1143;
Best Local Similarity 96.6%; Pred. No. 1.le-207;

Matches 1085; Conservative 2; Mismatches 33; Indels 3; Gaps 2;

QY 738 GGCAAGCGCGGAGCTCTACCGCATCCAGCATCCTCAACCCCATCAACATCTTCT 797
DB 21 GGCGCCCGGCGAGTATCTTAAAGATCCAGCATCCTCAACCCCATCAACATCTTCT 80
QY 798 GGTCAAGGTGCTGCTTTAGAGATCGTGACTCCGGGCCCAAGGTACCGGCAATCGGC 857
DB 81 GGTCAAGGTGCTGCTTTAGAGATCGTGACTCCGGGCCCAAGGTACCGGCAATCGGC 140
QY 858 CTTGACGCTCGCAACTTCTGTCCTGGCAGAGAGAGCTGAACAAAGTAGTGACAAGCA 917
DB 141 CTTGACGCTCGCAACTTCTGTCCTGGCAGAGAGAGCTGAACAAAGTAGTGACAAGCA 200
QY 918 CCCGAGTACTGGGACACTGCCATCTCTTACAGGAGGACTGTGTGGAGCCACAC 977
DB 201 CCCGAGTACTGGGACACTGCCATCTCTTACAGGAGGACTGTGTGGAGCCACAC 260
QY 978 CTGTGACACCTCGGCGATGCTGATGTGGTACCATGTGTGACCCCAAGAGAGCTGCTC 1037
DB 261 CTGTGACACCTCGGCGATGCTGATGTGGTACCATGTGTGACCCCAAGAGAGCTGCTC 320
QY 1038 TGTCAATGAGGAGATGGGCTTCCATCAGCTTCACCATGCCAGAGAGTGGGCCACGT 1097
DB 321 TGTCAATGAGGAGATGGGCTTCCATCAGCTTCACCATGCCAGAGAGTGGGCCACGT 380
QY 1098 GTTCAACATGCCCATGACAAATGTGAAAGTCTGTGAGGAGGTGTTGGGAAGCTCGAGC 1157
DB 381 GTTCAACATGCCCATGACAAATGTGAAAGTCTGTGAGGAGGTGTTGGGAAGCTCGAGC 440
QY 1158 CAACCATGATGTCCCGGACCTCATCCAGATCGACCGTGCACCCCTGTGACCGTG 1217
DB 441 CAACCATGATGTCCCGGACCTCATCCAGATCGACCGTGCACCCCTGTGACCGTG 500
QY 1218 CAGTGCTGCTCATCAGGAGTCTTCCAGAGCGGAGGAGTGTGCTCTGTGACCA 1277
DB 501 CAGTGCTGCTCATCAGGAGTCTTCCAGAGCGGAGGAGTGTGCTCTGTGACCA 560
QY 1278 ACCAGCAAGCCATCTCCCTGCGCGAGGATCTGCGCGGCGGAGTACACCTGAGCA 1337
DB 561 ACCAGCAAGCCATCTCCCTGCGCGAGGATCTGCGCGGCGGAGTACACCTGAGCA 620
QY 1338 GCAGTGGAGCTGGCTTTGGGTGGGCTCCAAAGCCCTGTCTTATACAGTACTGAC 1397
DB 621 GCAGTGGAGCTGGCTTTGGGTGGGCTCCAAAGCCCTGTCTTATACAGTACTGAC 680
QY 1398 CAAGCTGTGTGTCACCGGAAGCCCAAGGACAGATGTTGCCAGCCGCGCACTTCC 1457
DB 681 CAAGCTGTGTGTCACCGGAAGCCCAAGGACAGATGTTGCCAGCCGCGCACTTCC 740
QY 1458 CTGGGCGGATGGCAGCTGTGGCGAGGCGAAGCTCTGCTCAAGGGGCGCTGCTGGA 1517
DB 741 CTGGGCGGATGGCAGCTGTGGCGAGGCGAAGTCTGCTCAAGGGGCGCTGCTGGA 800
QY 1518 GAGACACACCTCAACAGCAGAGGTGATGTTCTTGGGCGAAATGGATCCCTATGG 1577
DB 801 AARACACACCTCAACAGCAGAGGTGATGTTCTTGGGCGAAATGGATCCCTATGG 860
QY 1578 CCCTGTCTGCGCACATGTGTGGGCGGTGCAAGTGGCGGAGGAGCAG-TGCACCAACC 1636
DB 861 CCCTGTCTGCGCACATGTGTGGGCGGTGCAAGTGGCGGAGGAGCAGTGCACCAACC 920
QY 1637 CCACCCC--TGCCAAAGGGGCGAAGTACTGCGAGGAGGTGAGGGTGAATACCGATCCTG 1694
DB 921 CCACCCCTGCCAACNGGGGCGAAGTACTGCGAGGAGGTGAGGGTGAATACCGATCCTG 980
QY 1695 CAACTGGAGCCCTGCCCCAGCTCAGCCTCGGGAAGAGCTTCGGGAGGAGCAGTGTGA 1754
DB 981 CAACTGGAGCCCTGCCCCAGCTCAGCCTCGGGAAGAGCTTCGGGAGGAGCAGTGTGA 1040
QY 1755 GCGTTTCAACGGGTACAACACAGCAGCAACCGGCTCACTCTCGCGGTGGCATGGTGC 1814
DB 1041 GCGTTTCAACGGGTACAACACAGCAGCAACCGGCTCACTCTCGCGGTGGCATGGTGC 1100

QY 1815 CAACTACTCCGCGTGTCTCCCGGACAAAGTGCAGCTCATC 1857
DB 1101 CAACTACTCCGCGTGTCTCCCGGACAAAGTGCAGCTCATC 1143

RESULT 9
AAD35568
ID AAD35568 standard; cDNA; 966 BP.
XX AAD35568;
XX
XX 26-JUL-2002 (first entry)
XX Human protease cDNA #1.
XX
XX Human; novel human, protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme; gene; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT 1..966
FT /*tag= a
FT /product= "Human protease #1"
XX
XX WO200226949-A2.
XX
XX 04-APR-2002.
XX
XX 27-SEP-2001; 2001WO-US30350.
XX
XX 29-SEP-2000; 2000US-236689P.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Friddle CJ, Hilbun E;
XX WPI; 2002-372123/40.
XX P-PSDB; AAE22540.
XX
XX Novel nucleic acid encoding a human protease, useful as a hybridization
XX probe for screening libraries and assessing gene expression patterns -
XX
XX Disclosure; Page 34; 41pp; English.
XX
XX The present sequence is a cDNA encoding novel human protein (NHP),
XX human protease. NHPs share structural similarity with animal proteases
XX particularly zinc metalloproteases. Sequences of the invention are
XX useful in therapeutic, diagnostic and pharmacogenomic applications.
XX NHP polynucleotides are used as hybridisation probes for screening
XX libraries and assessing gene expression patterns. They can also be
XX used for treating related biological disorders such as obesity, high
XX blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy.
XX
XX Sequence 966 BP; 161 A; 316 C; 305 G; 184 T; 0 other;
XX

Query Match 33.5%; Score 955.8; DB 24; Length 966;
Best Local Similarity 99.8%; Pred. No. 3.8e-189;
Matches 957; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCTTTTCTGGGATCCTTAACCTTGGCTTTCGCGGCGCAACCGCTGGAGGCTTCAG 60
DB 1 ATGCTTCTCTGGGATCCTTAACCTTGGCTTTCGCGGCGCAACCGCTGGAGGCTTCAG 60
QY 61 CCAGAGCGGAGGTAGTGTCTTCCATCGACTGGACCGGACATTAAGCGGCGCGCTAC 120
DB 61 CCAGAGCGGAGGTAGTGTCTTCCATCGACTGGACCGGACATTAAGCGGCGCGCTAC 120
QY 121 TACTGGGCGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTTCAGATCAACA-SCATTT 180
DB 121 TACTGGGCGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTTCAGATCAACA-SCATTT 180

Db 121 TACTGGGGGGTCCCGAGGACTCCGGGGATCAGGGGACTCATTTCATGATCAGACATTT 180
 QY 181 CAGGAGGACTTTTACCTACACCTGACCGGATGCTCAGTCTTCTGGCTCCCGCTTCTCC 240
 Db 181 CAGGAGGACTTTTACCTACACCTGACCGGATGCTCAGTCTTCTGGCTCCCGCTTCTCC 240
 QY 241 ACTGAGCATCTGGGCTCCCGCTCCAGGGGCTACCGGGGGCTTCTCAGACCTGCGAGC 300
 Db 241 ACTGAGCATCTGGGCTCCCGCTCCAGGGGCTACCGGGGGCTTCTCAGACCTGCGAGC 300
 QY 301 TGTCTTATCTGGGAGCTGAACCGGAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
 Db 301 TGTCTTATCTGGGAGCTGAACCGGAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
 QY 361 GGGGGGCTCCCGGAGCTTTGGCTACCGAGCGCGGAGCTATGCTATGATCCCGCTGCGC 420
 Db 361 GGGGGGCTCCCGGAGCTTTGGCTACCGAGCGCGGAGCTATGCTATGATCCCGCTGCGC 420
 QY 421 AATGCTAGCGCGCGGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 480
 Db 421 AATGCTAGCGCGCGGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 480
 QY 481 GGTGTTCCGGGCGGCTTCCCGGAGACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 540
 Db 481 GGTGTTCCGGGCGGCTTCCCGGAGACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 540
 QY 541 AACCCCGCCATCTACGGGCGCTTGGACCTTACAGCGCGGCGGCGGCGGCTTCGGGGAG 600
 Db 541 AACCCCGCCATCTACGGGCGCTTGGACCTTACAGCGCGGCGGCGGCGGCTTCGGGGAG 600
 QY 601 AGTCGTAGCGCGGAGCTTGGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 660
 Db 601 AGTCGTAGCGCGGAGCTTGGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 660
 QY 661 GAGACGCTGCTGCTCGGAGAGCTCAATGGTCAAGTTCACGCGCGGAGCGGAGCGGAGCG 720
 Db 661 GAGACGCTGCTGCTCGGAGAGCTCAATGGTCAAGTTCACGCGCGGAGCGGAGCGGAGCG 720
 QY 721 TATCTGCTAGCTGCTGCGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 780
 Db 721 TATCTGCTAGCTGCTGCGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 780
 QY 781 CCCATCAACATCTGCTGAGGCTGCTGCTTCTTAGAGATCGTCACTCCGGGCGGAGCGG 840
 Db 781 CCCATCAACATCTGCTGAGGCTGCTGCTTCTTAGAGATCGTCACTCCGGGCGGAGCGG 840
 QY 841 GTACCGGCAATCGGCGCTGACGCTGCGCACTTCTGCTGCTGCGAGAGAGAGCTGAGC 900
 Db 841 GTACCGGCAATCGGCGCTGACGCTGCGCACTTCTGCTGCTGCGAGAGAGAGCTGAGC 900
 QY 901 AAAGTGAGTGACAGCAACCGGAGTACTGGGAGACATGCGCATCTCTTACCGAGGAGGA 959
 Db 901 AAAGTGAGTGACAGCAACCGGAGTACTGGGAGACATGCGCATCTCTTACCGAGGAGGA 959

RESULT 10

AAH20226
 ID AAH20226 standard; DNA; 2670 BP.
 XX
 AC AAH20226;
 XX
 DT 07-AUG-2001 (first entry)
 XX
 DE Human ADAM-type metalloprotease-related probe, SEQ ID NO:3.
 DE
 KW Human: MDT54; ADAM-type metalloprotease; drug screening;
 KW A Disintegrin And Metalloprotease; cancer; arthritis; probe; ss.
 XX
 OS Homo sapiens.
 XX
 PN JP2001017183-A.
 XX
 PD 23-JAN-2001.

XX 09-JUL-1999; 99JP-0196584.
 PF 09-JUL-1999; 99JP-0196584.
 PR (YAMA) YAMANOUCHI PHARM CO LTD.
 XX WPI; 2001-275950/29.
 DR A new metal protease and its preparation for use as an anti-cancer and
 PT anti-arthritis therapeutic -
 XX Example 1; Page 11-12; 22pp; Japanese.
 XX The invention relates to the novel human ADAM (A Disintegrin And
 CC Metalloprotease)-type metalloproteases MDT54 (AAB73549) and MDT55
 CC (AAB73550). The metalloproteases can be used for the treatment of
 CC cancers and arthritis. The invention also relates to the genes encoding
 CC MDT54 and MDT55, vectors and host cells containing the MDT54 or MDT55
 CC genes, the recombinant production of MDT54 and MDT55, and antibody
 CC specific for MDT54 or MDT55, and methods of screening for compounds
 CC which modulate the activity of MDT54 and/or MDT55. The present
 CC sequence represents a human ADAM-type metalloprotease-related DNA
 CC sequence, used in an exemplification as a probe for screening human cDNA
 CC libraries in order to isolate MDT54 cDNA.
 XX
 SQ Sequence 2670 BP; 511 A; 853 C; 867 G; 439 T; 0 other;
 Query Match 23.3%; Score 666; DB 22; Length 2670;
 Best Local Similarity 59.2%; Pred. No. 7.2e-129;
 Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;
 QY 157 CTCATTTTTCAGATCAGACATTTTCAGGAGCTTTTACCTACACCTGACGCGCGGATGCT 216
 Db 157 CTCGCGCTCCACCTGCTCCGCTTCCGCAAGGGCTTCGCTGCTGCGCTGCGCGCGGAGC 216
 QY 217 CAGTTCCTTGGCTCCGCTTCTCCACTGAGCATCTGGGCGTCCCTCCAGGGCTCACC 276
 Db 217 AGCTTCCTTGGGCGCGGAGTTCAAGATCGAGCGCTTCGGGGGCTCCGGCGGGCG-ACC 273
 QY 277 GGGGCTCTTCAGACCTCGACGCTGCTTCTATTCTGGGAGCTGAACGCCGAGCGGAGC 336
 Db 274 GGGGCGAGCGGGGCTCGCGGCTGCTTCTTCTCCGSCACCGCTGAATGGGAGCGCGAG 333
 QY 337 TCGTTCGCTGCTGAGGCTGTGGGGGGGCTCCGCGAGGCTTTGGTACCGAGGCGCC 396
 Db 334 TCGTTCGCGGGGCTGAGGCTGTGGCGGGGCTGAGCGGCTCTTCTGCTGGAGCGGAG 393
 QY 397 GAGTATGTCATTAGCGCGCTGCCAATCTAGCGCGCGGGCGGCGAGCGCAACAGCGAG 456
 Db 394 GAGTTCACCATCCAGCGCGAGGCGCGGGGGCTCCCTGGCTCAGCGCGACCGCTCGAG 453
 QY 457 GCGGCACACCTTCTCAGCGCGGGGGTGTTCGGGGCGGGCTTCGGGAGACCCACCTCT 516
 Db 454 CGCTGGGCTCCGCGGAGCGCGCCCTCCCGCGAGGACCGGAGTGGGAGTGGAGAGC 513
 QY 517 CGCTCGGGGTG-CCCTCGGGGTGAACCCCGCATCTACGGGCGCTTGGACCTTACAA 575
 Db 514 GAGAGGCTCAGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 573
 QY 576 GCGCGCGCGGGGCTTCGGGGAGGAGTCTAGCGCGGCGAGG-TCCTGGGCGGCGC 630
 Db 574 GAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 633
 QY 631 AAGCGCTTTCGTCTATCCCGGCTACCTGGAGAGCGCTGGTGGTGGGAGCGAGTCAATG 690
 Db 634 AAGCGCTTTCGTCTAGGCGGCTTCTGAGAGCGCTGCTGGTGGTGGGCGAGTCCATG 693
 QY 691 GTCAAGTTCAGCGCGGAGCTGGAGCATTTATCTGCTGAGCGCTGCTGGCAACCGCGCG 750
 Db 694 GCTGCTTCTACGGGCGGAGCTGCAGAACACATCTCTGAGCTTAATGCTGTGGAGCG 753
 QY 751 CGACTCTACCGGCTCCCGAGCATCTCAACCCCATCATCTGTGTGTCAAGTGTGTG 810

Db	754	CGAATCTACAAGCACCCCGACATCAAGAATTCATCAACACTGTGGTGGTAAAAAGTCGTG	813
QY	811	CTTCTTGAAGATCGTACTCCGGGCCCAAGGTCAACGGCAATCGCGCCCTCAGCCTCGCG	870
Db	814	ATCTAGAGAATGAATAATGGGCCAGAGTGTCCGACATGGGGGCTTACACTCGCT	873
QY	871	AACTTCTGTGCTGGCAGAGAAGTGAACAAGTGAAGTGAACAAGCACCCCGAGTACTGG	930
Db	874	AACTTCTGCAACTGGCAGCGCGCTTTCACACGAGCCAGCAGCCGACCCAGAGCACTAC	933
QY	931	GACACTGCCATCTCTTCAACGAGGAGACCTGTGTGG---AGCCACCACTGTGACACC	987
Db	934	GACAGCGGCATCTGTCTCAACAGACAGAATTTCTGTGGCAGGAGGGCTGTGTGACACC	993
QY	988	CTGGGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAGAAGTGTCTGTTCATTGAG	1047
Db	994	CTGGGTGCGCAGACATCGGGACCATTTGTGACCCCAACAAAGTGTCTCGTGTATCGAG	1053
QY	1048	GACGATGGGTTCATATAGCCTTCACACTGCCACGAGCTGGGCCAGCTGTTCACAATG	1107
Db	1054	GATGAGGGCTCCAGCGGCCACACCTTGGCCCATGAATAGGACAGTCTCTCAGCATG	1113
QY	1108	CCCCATCACATGTGAAGTCTGTGAGGAGTGTTTGGGAAGCTCCGAGCCACACCATG	1167
Db	1114	CCCCAGCAGACTCCAAGCCCTGCACAGGCTTTCTGGGGCCCATGGGCACGACACAGTG	1173
QY	1168	ATGTCCCGGACCCCTCATCCAGATCGACCTGCCAACCCCTGGTCAGCTGCAGTGTGCC	1227
Db	1174	ATGCAACCGCTGTTCGTCCACCTGAACAGAGCTGCCCTGTCTCCCTCGACGGCCATG	1233
QY	1228	ATCATCACGACTTCCTGGACAGCGGGCAGGTGACTCCCTCTGGACCAACCCAGCAAG	1287
Db	1234	TATCTCACAGAGCTTCTGGACGGCGGCACGAGAGTGTCTCTGGATGCCCTGTCTGCG	1293
QY	1288	CCCATCTCCCTGCCGAGGATCTGCCGGC-----GCCAGCTACACCTTGAGCGCCAG	1341
Db	1294	GCCCTGCCCTCCCCACAGCCCTCCCGGCGCATGGGCCCTGTACCAAGCTGGACACAG	1353
QY	1342	TGCAGTGTGCTTTTGGCGTGGGCTCCAAGCCCTCTCCTTACA-----TGCAGTAC	1392
Db	1354	TGCAGGCAGATCTTTGGCGGGATTTCCGCCACTTCCCCACACCTCTGCTCAGGAGCTC	1413
QY	1393	TGCACCAAGCTGTGTG---CACCGGAGGCGCAAGGACAGATGTGTGCCAGACCCCGC	1449
Db	1414	TGCGCCCAAGCTTTGTGTCCACACTGATGGGCTGAGCCCTGTGCCACACGAAGATGGC	1473
QY	1450	CACCTCCCTGGGCCATGGCACCAGCTGTGGCGAGGCGCAGCTCTGCTCTAAAGGGGCC	1509
Db	1474	AGCTTGCCCTGGGCTGACGGCAGCGCTGGCGGCGCTGGGCACCTCTCTCAGAAGGCGC	1533
QY	1510	TG-----CGTGGAGAGACACACCTCAACAGACAGAGGTGGATGTTCTCTGGGCCAA	1563
Db	1534	TGCTACTTGAGGAGAAGTGGAGAGGCCAAGCCCTGTGTAGATGAGGCTGGGCACCG	1593
QY	1564	TGGGATCCCTATGGCCCTCTCTCGGCACATGTGTGGGGCGCTGCAGCTGGCCAGGAGG	1623
Db	1594	TGGGACCTCTGGGGAGAAATTTCTCGGACCTGTGGAGGAGGATACAGTTTTCACACCGT	1653
QY	1624	CAGTGCACCAACCCACCCCTGCCAACGGGGCGCAAGTACTCGGAGGAGTGTAGGGTGA	1683
Db	1654	GAGTGCAAAGACCCGAGCCCTCAGAATGGAGGAAGATACTGCTTGGCTCGGAGAGCAAG	1713
QY	1684	TACCGATCTTGCAACTTGAGCCCTTGCCCAAGCTCAGCCTCCGGAAGAGCTTCCGGAG	1743
Db	1714	TACCACTATGCCACAGGAGGAATGCC-----CCTGACGGGAAGAGCTTCAGGGAG	1767
QY	1744	GACAGTGTGAGCTTTCAACGGCTTACAACACAGCAGCAACACCGCTCTCACTCGCGGTG	1803
Db	1768	CAGCAGTGTGAGAAGTATAATGCCCTACAATTTACACTGACATGGACGGGAATCT---CCTG	1824
QY	1804	GCATGGTGGCCAAAGTACTCCGGCTGTCTCCCGGGACAAAGTGCAGGCTCATCTGCCGA	1863

Db	1825	CAGTGGGTCCCAAGATATGCTGGGGTGTCC	CCCCGGGACCGCTGCAAGTGTCTTC	GC	CGGA	1884					
Qy	1864	GCAATGCACTGGCTACTTCTATGCTGGC	ACCAAGGTGGTGGAGGCACTGTGC	1923							
Db	1885	GC	CGGGGAGGAGCGAGTTC	AAAGTGTTCAGGCCAAGGTGATGATGGT	ACCCTGTGT	1944					
Qy	1924	TCTCTGACTCCACTCCGTCTGTGTCAAG	AGCAAGTGCATCAAGGCTGCTGTGATGG	1983							
Db	1945	GGGCCAAGAACACTGGCCATCTGTGTCG	TGGCCACTGTCTCAAGGCCGTGTC	2004							
Qy	1984	AACTGGGCTCCAAGAGAGATTCGACAAG	TGTGGGTGTGGGGGAGACAAATAAGAC	2043							
Db	2005	GTGTGTGACTCGCCTCGAAGCTGGACA	ATATGCGGGTGTGTGGGGCAAGGCAACTCC	2064							
Qy	2044	TGCAAGAGGTGACTGTGACTCTTCAACC	AGCCCATCATGGCTACAAATTCGTGTGGCC	2103							
Db	2065	TGCAAGAGGTCTCCGGTCCCTCACCC	CCCAATATTTGGTACAATGACATTTGTCA	2124							
Qy	2104	ATCCCGCAGGCGCTCAAGCATCGACAT	CCGGCAGCGGTTTACAAAGGGCTGATCGG	2163							
Db	2125	ATCCAGCTGTGCGCACTAATATTGAC	TGTAAGCAGCGGAGCCACCCGGGTGTGC	2184							
Qy	2164	GATGACAACACTACCTGGCTGTGAAG	ACAGCAAGCAAGCAAGTACTGCTCAACGGG	2223							
Db	2185	GATGGAACTACCTGGCGCTCAAGAG	CGGCTGATGGCAGTACTGCTCAACGGCA	2244							
Qy	2224	GTGTGTGCGGCTGGAGCGGACCTGGT	GTGAAGGCGAGTCTGCTGGGTACAGCGGC	2283							
Db	2245	GCAATCTCTGCATAGAGAGGACATCT	TGGTGAAGGGAGCCATCCTGAAGTACAG	2304							
Qy	2284	ACGGSCACAGCGGTGGAGAGCCTTC	CAGGCTTCCCGGCCCATCCTGGAGCGCTG	2343							
Db	2305	TCCATCGCCACCTGGAGCGCCTCG	AGACTTCCGGCCCTTCCAGAGCCTCTG	2364							
Qy	2344	GAGTCTCTCCGT---GGGAGATGAC	ACACCGCCCCGGGTCCGTACTCTCTATCTG	2400							
Db	2365	CAGCTCTGACAGTCCCTGGCGAG	GTCTTCCCCCAAAGTCAAAATACACCTCT	2424							
Qy	2401	CCCAAGA	2408								
Db	2425	CCTAATGA	2432								
RESULT 11											
ID	AAF82166 standard; cDNA; 2670 BP.										
XX	AAF82166;										
AC											
DT	02-JUL-2001 (first entry)										
XX	Human ADAM type metal protease MDT53 encoding cDNA SEQ ID NO:21.										
DE											
XX	Human; a disintegrin and metalloprotease type metal protease; MDT51;										
KW	MDT52; MDT53; ADAM type metal protease; cytosolic; antiarthritic;										
KW	cancer; arthritis; arthrosis deformans; ss.										
XX											
OS	Homo sapiens.										
XX											
FH	Key	Location/Qualifiers									
FT	CDS	1..2670									
FT		/*tag= a									
FT		/product= "MDT53"									
FT		/note= "a disintegrin and metalloprotease (ADAM)									
FT		type metal protease"									
XX											
PN	JP2001008687-A.										
XX											
PD	16-JAN-2001.										
XX											
PF	25-JUN-1999; 99JP-0180973.										
XX											
PR	25-JUN-1999; 99JP-0180973.										

Qy 1984 AACCTGGCTCCAAAGAGATTGCAAGTGTGGGTGTGGGGAGACAATAAGACG 2043
 Db 2005 GTGTGGACTCGCTCGAAGCTGGACAATGCGGGTGTGGGGCAAGGCAACTCC 2064
 Qy 2044 TGCAGAGGTGACTGAGCTCTTACCAAGCCCATGATGGCTACAAATTTCTGGTGGCC 2103
 Db 2065 TGCAGAGGTGCTCGGGTCCCTCACCCCAATATGCTTACAAATGACATTTGTCACC 2124
 Qy 2104 ATCCCCCAGCGGCTCAAGCATCGACATCCGCGGCTTACAAAGGCTGATCGGG 2163
 Db 2125 ATCCCACTGGTCCCAATATTTGAGTGAAGCAGCGGACCCCGGTGTGCAGAAC 2184
 Qy 2164 GATGACAACCTACCTGGCTCTGAAGACAGACCAAGCAAGTACCTGTCAACGGGCAATTC 2223
 Db 2185 GATGGAACTACCTGGCGCTGAAGCGGCTGATGGCAGTACTGCTCAAGGCAACCTG 2244
 Qy 2244 GTGGTGTGGGGTGGAGCGGACCTGGTGTGAAGGCAAGTGTGTCGGGTACAGCGG 2283
 Db 2245 GCCATCTCTGCTATAGAGAGGACATCTTGGTGAAGGGGACCATCTCTGAAGTACAGCGG 2304
 Qy 2284 ACGGGACAGCGGTGGAGAGCTTGCAGGCTTCCCGGCCATCTGGAGCGGCTGACCGG 2343
 Db 2305 TCCATGCCACCTGGAGCGCTGCAGAGCTTCCGGCCCTTGCAGAGCTCTGACAGTG 2364
 Qy 2344 GAGTCTCTCCGT---GGGGAAGATGACACCGCCCGGTCGGTCTCTTATCTG 2400
 Db 2365 CAGCTCTGACAGTCCCTGGCGAGGTCTCCGCCCAAAAGTCAATACACCTTCTTTGTT 2424
 Qy 2401 CCCAAGA 2408
 Db 2425 CCTAAGA 2432

RESULT 12

AAZ32001

ID AAZ32001 standard; cDNA; 3008 BP.

XX AC AAZ32001;

XX DT 10-JAN-2000 (first entry)

XX DE Human METH2 encoding cDNA.

XX KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;

KW cancer; diagnosis: hyperproliferative disorder; autoimmune disease;
 KW angiogenesis inhibitor; abnormal wound healing; inflammation;
 KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
 KW diabetic retinopathy; macula degeneration; haemangioma; detection;
 KW arterial-venous malformation; immune deficiency; ss.

XX OS Homo sapiens.

XX XX WO937660-A1.

XX PD 29-JUL-1999.

XX XX 22-JAN-1999; 99WO-US01313.

XX XX 23-JAN-1998; 98US-0072298.

XX PR 28-AUG-1998; 98US-0098539.

XX XX (IRUE/) IRUELA-ARISPE L.

XX PA (HAST/) HASTINGS G A.

XX XX (RUBE/) RUBEN S M.

XX PI IrueLa-Arispe L, Hastings GA, Ruben SM;

XX DR WPI; 1999-590684/50.

XX DR P-PSDB; AAY49502.

XX XX New isolated metalloprotease thrombospondin polypeptides, useful for

XX PT treating hyperproliferative disorders, cancers or autoimmune disorders

XX PT

XX

PS Claim 4; Fig 2; 457pp; English.

XX

AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human metalloprotease thrombospondin (METH) proteins METH1 and METH2 respectively. METH1 and METH2 have been found to be potent inhibitors of angiogenesis both in vitro and in vivo. They can be used for treating cancer and other disorders related to angiogenesis including abnormal wound healing, inflammation, rheumatoid arthritis, psoriasis, endometrial bleeding disorders, diabetic retinopathy, some forms of macula degeneration, haemangiomas, and arterial-venous malformations. They may be useful in treating deficiencies or disorders of the immune system by activating or inhibiting the proliferation, differentiation, or mobilisation (chemotaxis) of immune cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or toxins), or infectious. They can also be used to treat inflammatory conditions, both chronic and acute conditions. The products can also be used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to AAY49511 represent sequences given in the exemplification of the present invention.

XX Sequence 3008 BP; 617 A; 924 C; 948 G; 514 T; 5 other;
 SQ

Query Match 23.3%; Score 666; DB 20; Length 3008;

Best Local Similarity 59.2%; Pred. No. 7.3e-129;

Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;

Qy 157 CTCATTTTTCAGATCACAGCATTTTCAGGAGGACTTTTACCTACACCTGAMGCCCGATGCT 216
 Db 160 CTCGGCTCCACCTGTCCGCTTCGGCAAGGCTTCGNGTTGGCTCGCGCCGACGAC 219
 Qy 217 CAGTCTTGGTCCCGCTTCTCCACTGAGCATCTGGCGGTCCCGCTCCAGGGGTCAAC 276
 Db 220 AGCTTCTTGGCGCGGCTTCAAGATCGAGCGCTTCGGGGGCTCGCGCGGGG---ACC 276
 Qy 277 GGGGCTCTTCAGACCTCGGACGCTGTCTTATTCTGGGACGTGAAGCCGACGCGGAC 336
 Db 277 GGGGCGAGCGGGGCTCGCGGCTGTTTTTTCGGCACCTGATGGGAGCCGCGAG 336
 Qy 337 TCGTTCGTGTGTGAGCCTGTGCGGGGGGCTCCGCGAGGCTTTTGGCTACCGAGGCGCC 396
 Db 337 TCGTTCGTGTGTGAGCCTGTGCGGGGCTGAGCGGCTCTCTCTCTGCGAGCGGAG 396
 Qy 397 GAGTATGTCAATAGCCCTGCCAATCTAGCGCGCGCGGCGGCGAGCGCAAGACGACG 456
 Db 397 GAGTACCATCCAGCCGCGGCGGGGGGCTCCCTGGCTCAGCGCGCACCGCTCGAC 456
 Qy 457 GCGGCACACCTTCTCCAGCGCGGGGGTGTTCGGGGCGGGGCTTCCGAGAGACCCACCTCT 516
 Db 457 CGCTGGGTCCCGCGGAGCCGCCCTCCCGCGAGSACCCGAGTGGAGTGGAGACG 516
 Qy 517 CGCTCGGGGTG-GCCTCGGCTGGAAACCCGCCATCTACGGGCCCTTGACCCCTTACAA 575
 Db 517 GGAGAGGTTCAGAGGAGGAGAGAGAGAGACCAACAGGAGGACAGCAGGAGGAGAGCAA 576
 Qy 576 GCCGCGCGGGGCGCTTCGGGGGAGAGTCTGAGCGCGCAGG-----TCTGGCGGCGCC 630
 Db 577 GAAGAGGAGGAGAGGCGCTAGCGGCCGCCACCGCCCTGGGGGCGCAGAGTAGAGACC 636
 Qy 631 AAGCGTTTCGTGTATATCCCGCGGTACGTGGAGACGCTGTGTGTCGGGACGATCAATG 690
 Db 637 AAGCGTTTCGTGTATATCCCGCGGTTCGTGGAGACGCTGTGTGTCGGGCGATGCGTCCATG 696
 Qy 691 GTCAGTTCCAGCGCGGACCTGGAACATATCTGCTGAGCTGTGCTGGCAACGCGGCGG 750
 Db 697 GCTGCCCTTCTACGGGGCGACCTGCAGAACACATCTCGAGCTTAATGTCTGTGCGAGCC 756
 Qy 751 CGACTCTACCGCCCATCTCCAGCATCTCAACCCCATCAACATCGTTGTGTCAAGTGTGCTG 810
 Db 757 CGATCTACAAGACACCCAGCATCAAGATTCATCAACCTGATGGTGGTAAATGTGCTG 816
 Qy 811 CTTCTTAGAGATCGTGACTCCGGGCCCAAGGTACACCGCAATGCGGCCCTGACGCTGCGC 870

Db 817 ATCGTAGAAGTGAATAATGGGCCAGAGGTGTCCGACAATGGGGGCTTACATGCGT 876
Qy 871 AACTTCTGTGCTGGGAGAGAGCTGAACAAGTGAAGTACAGACACCCGAGTACTGG 930
Db 877 AACTTCTGCAACTGGCAGCGGCTTTCAACAGCCAGCGACGCCACCCAGAGCACTAC 936
Qy 931 GACACTGCCATCTCTTCAACAGCGAGGACCTGTGTGG---AGCCACCACCTGTGACACC 987
Db 937 GACAGGCCATCTCTTCAACAGACAGAACTTCTGTGGCAGGAGGGGTGTGTGACACC 996
Qy 988 CTGGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAGAGTGTCTGTGTAATTG 1047
Db 997 CTGGGTGTGCAGACATCGGACCAATTTGTGACCCCAACAAGTGTCTCGGTGATCGAG 1056
Qy 1048 GAGATGGGCTTCCATCAGCCTTACACATGTCACAGTGGCCAGAGTGGCCAGCTGTTCACATG 1107
Db 1057 GATGAGGGGCTCAGGGGGGCCACACCTTGGCCCATGAATAGGGCACCTCTCTCAGCATG 1116
Qy 1108 CCCCATCACAATGTGAAGTCTGTGAGGAGTGTGTGGGAAGCTCCGAGGCAACACATG 1167
Db 1117 CCCACAGACATCCAGCCCTGCACAGGCTCTTGGGGCCATGGGCAAGCACACGTG 1176
Qy 1168 ATGTCCCGACCTTATPACAGATCGACCGTGTCCAAACCCCTGTGACGCTGCACTGTGCC 1227
Db 1177 ATGSCACCGCTGTCTGTCACCTGAACACAGACGCTGCCCTGTGTCCCTCGACGCGCATG 1236
Qy 1228 ATCATACCCAGTCTCTGGACAGGGGACGCTGACTGCTCTGTGACCAACCCACGAG 1287
Db 1237 TATCTCAGAGCTTCTTGGACGGGGGACGAGAGCTGTCTCTGTGATGGCCCTGTGTGGG 1296
Qy 1288 CCCATCTCCCTGCCCGAGGATCTGCCGGG-----GCCAGCTACACCTCTGACCCAGCAG 1341
Db 1297 GCCCTGCCCTCCACAGGCTCCCGGGCGGCATGSCCTGTACCAGCTGGACACGAGCAG 1356
Qy 1342 TGCAGTGGCTTTTGGCGTGGGCTCAAGCCCTGTCTTACA-----TGCAGTAC 1392
Db 1357 TGCAGGAGAGATCTTTTGGCGCGGATTTCCGCCACTGTCCGCCCAACACACCTCTGTCTCAGGACGTC 1416
Qy 1393 TGCACCAAGCTGTGGTG---CACCGGAAGGCCAAGGACAGATGTGTGCCAGACCCGC 1449
Db 1417 TGCGCCAGCTTTGTGCTCCACACTGTGAGGCTGAGGCTGAGGCTGGCCAGATGCG 1476
Qy 1450 CACTTCCCTGGCCGATGGACAGCTGTGGGAGGGGCAAGCTCTGCCCTCAAGGGGCC 1509
Db 1477 AGCCTGGCTGGCTGACGGCAGCCCGCTGGGGCTGGGACCTCTGCTCAGAGGCGAGC 1536
Qy 1510 TG-----CGTGGAGAGACAACTCAACAAGCACAGGCTGGATGTTCTTGGGCCAAA 1563
Db 1537 TGTCTACTGAGGAGGAAGTGGAGGCCCAAGCCCTGTGTAGATGAGGCTGGCCACCG 1596
Qy 1564 TGGGATCCCTATGGCCCTGTCTCGGCACATGTGTGGGGCGCTGACGCTGGCCAGGAGG 1623
Db 1597 TGGGACCTGGGAGATGTTCTCGGACCTGTGGAGGAGGTACAGTTTTCACACCGT 1656
Qy 1624 CAGTGCACCAACCCCTTCCCAAGGGGCAAGTACTGCGAGGAGTGAAGGTGAAA 1683
Db 1657 GAGTGCAGAGGCCCGAGCTCAGATGGAGGAATATGCTTGGGTGGGAGGAGCAAG 1716
Qy 1684 TACCGATCTCAGACCTGGAGCCCTGCGCCAGCTCAGCCTCCGGAAAGAGCTTCCGGGAG 1743
Db 1717 TACCAGTATGCCACAGGAGGAATGCC-----CTGACGGGAAAGACTTCAGGGAG 1770
Qy 1744 GAGCAGTGTGAGGCTTTCAACGGCTACAAACACAGCACCACCGGCTCACTCTGGCGGTG 1803
Db 1771 CAGCAGTGTGAGAGATATATGCTACAAATTAACACTGACATGGACGGGAATCT---CCTG 1827
Qy 1804 GCATGGGTGCCAAGTACTCGGCGTGTCTCCCGGGACAGTGTCAAGCTCATCTGCCGA 1863
Db 1828 CAGTGGGTCCCCAAGATGTCTGGGGTGTCCCGCCGGAGCCGCTCAAGTGTCTGCCGA 1887
Qy 1864 GCAATGCACTGGCTACTTCTATGTGTGTCACCCCAAGTGTGTGGACGGCAGCCTGTGC 1923

Db 1888 GCGCGGGGAGGAGGAGCTTCAAAAGTGTTCAGGCGCAAGGTGATGATGGGACCTGTGT 1947
Qy 1924 TCTCTGTACTCCACTCCGCTGTGTCCAGGCAAGTGCATCAAGGCTGGCTGTGATGGG 1983
Db 1948 GGGCCAGAACACATGGCCATCTGTTCGTGCGTGCAGTGTGTCAAGGCGGCTGTGACCAT 2007
Qy 1984 AACCTGGGCTCCAAGAGAGATTCGACAAGTGTGGGTGTGGGGGAGACAATAAGAGC 2043
Db 2008 GTGTGGTACTCGCTCGGAAGCTGGACAATCGGGGTGTGTGGGGGCAAGAGCAACTCC 2067
Qy 2044 TCCAGAGAGGTGACTGGACTCTTCCACCAAGCCCATGATGCTCAATTTCTGTGTGGCC 2103
Db 2068 TSCAGGAAGGTCTCCGGGTCCCTCACCCACCAATTAATGCTCAATGACATTTGTACC 2127
Qy 2104 ATCCCGCAGGGCGCTCAAGCATCGACATCCGCGCAGCGGCTTACAAAGGCTGTATCGG 2163
Db 2128 ATCCAGCTGTGGCCACTAATATTGACCTGAAGCAGCGGAGCCACCGGGTGTGCAGAAC 2187
Qy 2164 GATGACAACCTACCTGGCTCTGAAGAACAGCAAGCAAGTACCTGCTCAACGGGCAATTTC 2223
Db 2188 GATGGAACTACTCTGGCTGAAGACGGCTGATGGGCGACTACCTGCTCAACGGCAACCTG 2247
Qy 2224 GTGGTCTCGGCGTGAAGGACCTGTGTGAAGGCGAGTCTGCTGCGGTACAGCGGC 2283
Db 2248 GCATCTCTGCCATAGAGGACATCTTGTGAAGGGGACCATCTCTGAAGTACAGCGGC 2307
Qy 2284 ACGGCACAGCGGTGGAGAGCCTCGAGGCTTCCCGGCCCATCTCTGGAGCCGCTGACCGTG 2343
Db 2308 TCCATCGCACCTCGAGCGCTGCAGAGCTTCCGGCCCTTGCCAGAGCCTCTGCACAGTG 2367
Qy 2344 GAGGTCTCTCGT---GGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTATCTG 2400
Db 2368 CAGCTCTCTGACAGTCCCTGGCGAGGCTTCCCGCCCAAAAGTCAATACACCTTCTTGT 2427
Qy 2401 CCCAAGA 2408
Db 2428 CCTAATGA 2435
RESULT 13
AAC90058
ID AAC90058 standard; DNA; 3008 BP.
XX AAC90058;
AC AAC90058;
XX 19-MAR-2001 (first entry)
XX Human METH2 coding sequence.
XX Human; METH2; metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiofibroma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control; ss.
XX Homo sapiens.
OS WO200071577-A1.
XX 30-NOV-2000.
XX 25-MAY-2000; 2000WO-US14462.
XX 25-MAY-1999; 99US-0318208.
PR 20-JUL-1999; 99US-0144882.
PR 10-AUG-1999; 99US-0147823.
PR 13-AUG-1999; 99US-0373658.
PR 22-DEC-1999; 99US-0171503.


```

Db 1657 GAGTGCAGAGCCCGGAGCTCAGATGGAGGAGATAGTCCCTGGTGGAGAGCCAAAG 1716
Qy 1684 TACCGATCTGCAACCTGGAGCCCTGCCAGCTCAGCTCCCGGAAAGAGCTTCCGGGAG 1743
Db 1717 TACCACTGATGCAACAGGAGGAGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1770
Qy 1744 GAGCAGTGTGAGGCTTCAACGGCTCAACACAGCAGCAGCAGCAGCAGCAGCAGCAG 1803
Db 1771 CAGCAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1827
Qy 1804 GCATGGTGGCCCAAGTACTCCGGGCTGTCTCCCGGAGCAGCAGCAGCAGCAGCAGCAG 1863
Db 1828 CAGTGGTGGCCCAAGTACTCCGGGCTGTCTCCCGGAGCAGCAGCAGCAGCAGCAGCAG 1887
Qy 1864 GCAATGGCAGTGGCTTCTATGCTGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1923
Db 1888 GCGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1947
Qy 1924 TCTCTGATCCACCTCCGCTGTGTCCAAAGGCAAGTGCATCAAGGCTGGCTGTGATGGG 1983
Db 1948 GGGCAGAAACACTGGCCATCTGTCTCCGCTGGCAGTGTCTCAAGGCGGCTGTGACCAT 2007
Qy 1984 AACCTGGGCTCCAAAGAGATTCGCAAGTGTGGGCTGTGTGGGAGGAGCAATAGAGC 2043
Db 2008 GTGGTGGACTCGGCTCGGAAGCTGGACAAATGGGGGTGTGTGGGGCAAGCAACTCC 2067
Qy 2044 TCGAAGAGTGTGACTGACTCTTCCACCAAGCCATGATGCTACATTTCTGTGGTGGCC 2103
Db 2068 TCGAAGAGTGTGACTGACTCTTCCACCAAGCCATGATGCTACATTTCTGTGGTGGCC 2127
Qy 2104 ATCCCGCAGCGGCTCAAGCATCGACATCCGCGAGCGGCTTACAAAGGGCTGATGGG 2163
Db 2128 ATCCCGCAGCGGCTCAAGCATCGACATCCGCGAGCGGCTTACAAAGGGCTGATGGG 2187
Qy 2164 CATCAACACTACCTGCTCTGAAGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2223
Db 2188 GATGGGAACACTACCTGCTCTGAAGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2247
Qy 2224 GTGGTGTGGGCGGTGGAGCGGAGCTGTGTGAAGGCGAGTCTGTCTGCGGTACAGCGC 2283
Db 2248 GCCATCTGCGCATAGACGAGGACATCTGTGTGAAGGCGAGTCTGTCTGCGGTACAGCGC 2307
Qy 2284 ACGGCGCAGCGGTGGAGAGCTCGAGCTTCCGCGGCGATCTCGGAGCGGCTGACCGTG 2343
Db 2308 TCCATCGCCACCTGGGCGCTCGAGAGCTTCCGCGGCGATCTCGGAGCGGCTGACAGTG 2367
Qy 2344 GAGTCTCTCCGT---GGGGAAGATGACACCGCGCGGTGCTACTCTCTATCTG 2400
Db 2368 CAGCTCTGACAGTCCCTGGCGAGGCTTCTCCCGGAGGAGTCTCTCCCGGAGGAGTCTCT 2427
Qy 2401 CCCAAGA 2408
Db 2428 CCTAATGA 2435

```

RESULT 14

AAZ32010
ID AAZ32010 standard; DNA; 4180 BP.

XX

AC AAZ32010;

XX

DT 10-JAN-2000 (first entry)

XX

DE Human METH1 related EST D67076.

XX

KW Human; METH1; METH2; anti-angiogenic; metalloproteinase thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
KW arterial-venous malformation; immune deficiency; ss.

XX

OS Homo sapiens.
PN WO9937660-A1.
XX
PD 29-JUL-1999.
XX
PF 22-JAN-1999; 99WO-US01313.
XX
PR 23-JAN-1998; 98US-0072298.
XX
PR 28-AUG-1998; 98US-0098539.
XX
PA (IRUELA) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
XX
PI IrueLA-Arispe L, Hastings GA, Ruben SM;
XX
XX WPI; 1999-590684/50.
XX
XX New isolated metalloproteinase thrombospondin polypeptides, useful for
XX treating hyperproliferative disorders, cancers or autoimmune disorders
XX
XX Disclosure; Page 243-246; 457pp; English.
XX
XX AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
XX metalloproteinase thrombospondin (METH) proteins METH1 and METH2
XX respectively. METH1 and METH2 have been found to be potent inhibitors of
XX angiogenesis both in vitro and in vivo. They can be used for treating
XX cancer and other disorders related to angiogenesis including abnormal
XX wound healing, inflammation, rheumatoid arthritis, psoriasis,
XX endometrial bleeding disorders, diabetic retinopathy, some forms of
XX macula degeneration, haemangiomas, and arterial-venous malformations.
XX They may be useful in treating deficiencies or disorders of the immune
XX system, by activating or inhibiting the proliferation, differentiation,
XX or mobilisation (chemotaxis) of immune cells. The etiology of these
XX immune deficiencies or disorders may be genetic, somatic, such as
XX cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
XX toxins), or infectious. They can also be used to treat inflammatory
XX conditions, both chronic and acute conditions. The products can also be
XX used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to
XX AAY49511 represent sequences given in the exemplification of the present
XX invention.
XX
XX Sequence 4180 BP; 1078 A; 1051 C; 1146 G; 905 T; 0 other;

Query Match 22.3%; Score 636; DB 20; Length 4180;

Best Local Similarity 60.7%; Pred. No. 1.3e-122;

Matches 1098; Conservative 0; Mismatches 700; Indels 12; Gaps 3;

Qy 631 AACGCTTTGCTGCTATCCCGGCTAGCTGGAGACGCTGCTGCGGAGCAGTCAATG 690
Db 1143 AAGCGATTTGTGTCCAGCCCCCGTTATGTGGAACCATCTCTGTAGTGACAGTCCATG 1202
Qy 691 GTCAAGTTCACGCGCGGACCTGGAACATTATCTGCTGACGCTGCTGGCAACGCGGG 750
Db 1203 GCCGACTTCCACGCGAGCGTCTAAAGCATTTACCTTCTAACCTGTTCTCGTGCGAGCC 1262
Qy 751 CGACTCTACGGCCATCCGAGCATCTCAACCCCATCAACATGTTGTGTGTCAGGTGCTG 810
Db 1263 AGGTTTACAGCATCTCCAGCATTTAGGAATTCATTTAGCTGTTGTGTGTAAGATCTTG 1322
Qy 811 CTTCTTAGAGATCGTGACTCCGGGCGCAAGGTCACCGCAATGCGGCGCTGACGCTGCC 870
Db 1323 GTCATATACGAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1382
Qy 871 AACTTCTGTGCTCGGAGAGAGCTGAACAAAGTGAAGTGAACAGACCCCGAGTACTCG 930
Db 1383 AATTTCTGCGAGCTGCGAGAAACAACAACAGCCCGAGTGAAGGAGTCCAGAGCACTAT 1442
Qy 931 GACACTGCCATCTCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 990
Db 1443 GACACTGCAATTTCTGTTCCAGCAGAGGATTTATGTGGGCTCCACACGTTGACACTCTC 1502

Qy	991	GGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAGAAAGTCTCTCTGTCAATGAGGAC	105
Db	1503	GAATGGCGAGATGTTGAAACCGTATGTGACCCACGAGAGCTGCTCACTATAGAAGAT	1562
Qy	1051	GATGGCTTCCATCAGCCTTTCACCACTGCCCCACGAGCTGGCGACAGTGTTCACATGCC	1110
Db	1563	GATGGTTTGAAGCGCGCTTCACACAGCCCATGAATTGGGCCATGTGTTTAAACATCGC	1622
Qy	1111	CATCACAATGTGAAGTCTGTGAGGAGTGTGTTGGAAAGCTCCGAGCCCAACCATGATG	1170
Db	1623	CACGATGATGCTAAGCACTGTGCCAGCTTGAATGTGTGAGTGGCGATTTCTCATCTGATG	1682
Qy	1171	TCCCCGACCTCATCCAGATCGACCGTGCACACCCCTGTGTCAGCTGCAGTGCCTGCCATC	1230
Db	1683	GCCTCGATGCTCTCCAGCTTAGACCATAGCCAGCCCTGTCACCTTGCAGTGCCTACATG	1742
Qy	1231	ATCACCGACTTCTTGACACGCGGCACGTGACTGCCTCTGGACCAACCCAGCAAGCCC	1290
Db	1743	GTCACGTCTCTTAGATAATGACACGCGGAATGTTGTAGGACAAGCCCCAGAAATCCA	1802
Qy	1291	ATCTCCTCTGCCGAGGATCTCGCGGCGCCAGCTACACCCCTGAGCCAGCAGTGGAGCTG	1350
Db	1803	ATCAAGCTCCCTTCTGATCTTCCCGTACTTTGTACGATGCCAACCGCCAGTGTCACTTT	1862
Qy	1351	GCTTTTGGCTGGGCTCCAAAGCCCTGCTTACATGCAGTAC---TGCACCAAGCTGTGG	1407
Db	1863	ACATTCGAGAGAAATCCAGCACTGCCCTGATGCAGCCAGCAATGTACTACCCCTGTGG	1922
Qy	1408	TGCACCGGAAGGCCAAGGACAGATGTGTGCCAGACCCGCCACTTCCCTGGGCGCGAT	1467
Db	1923	TGCACTGGCACCTCCGCTGGCTTACTGTGTGTGCCAAACAAACAACTTCCCTTGGCGAGAT	1982
Qy	1468	GGCACAGCTGTGGCAGGGCAGCTCTGCCCTCAAGGGGCGCTGGTGGAGAGACACAAC	1527
Db	1983	GGCACAGCTGTGGGAAGGGAAAGTGGTGTGTCACTGGCCTGCGTGAACAAGACAGAC	2042
Qy	1528	CTCAACAAGCACAG-----GGTGTGATGTTCTCTGGGCCAAATGSGATTCCTATGCGCCC	1581
Db	2043	ATGAAGCATTTTGCTACTCTCTTTCATGGAAGCTGGGACCACTGGGAGACCTGGGGAGAC	2102
Qy	1582	TGCTCGCGCACATGTGGTGGGGCGGTGCAGCTGGCCAGGAGGCAGTGCACCAACCCCAAC	1641
Db	2103	TGCTCAAGAACTGTGGTGGTGGAGTTCAATACACAATGAGAGAATGTGACAACCCAGTC	2162
Qy	1642	CCTGCCAACGGGGCAAGTACTCCGAGGAGTGAAGGTGAATACCGATCCTGCAACCTG	1701
Db	2163	CCAAAGACGGGGAAGTACTGTGAAGGCCAAACAGTCCGCTACAGCTCCTGTAAACATC	2222
Qy	1702	GAGCCCTTGCCCGAGCTCAGCCTCCGGAAGAGCTTCCGGGAGGAGCAGTGTGAGGCGTTTC	1761
Db	2223	GAGGACTGTCCAGACAATAA---CGGAAAACGTTACAGAGGAGCAGTGGGAGGCGCAC	2279
Qy	1762	AAGGGCTACACACAGCACACCACCGGCTCACTTCGCGCTGGCATGGGTGCCCAAGTAC	1821
Db	2280	AATGAGTTTTTCCAAAGCTTCTTTTGGGAATGAGCCACTGTAGAGTGGACACCAAGTAC	2339
Qy	1822	TCCGGGGTGTCTCCCGGGACAAAGTGCAAGCTCATCTCCGAGGCCAATGGACTGGCTAC	1881
Db	2340	GCCGGGGTCTCGGCAAGGACAGTGTGCAAGCTCACTGTGAAGCCAAAGGCAATGGGTAC	2399
Qy	1882	TTCTATGTGTGCGACCCAAAGTGGTGGACGSCAGCTGTGCTCTCTGTACTCCACCTCC	1941
Db	2400	TTTTTCGTCTTACAGCCCAAGTTGTAGATGGCACTCCCTGTACTGCAGACTCTACCTCT	2459
Qy	1942	GTCTGTGTCCAGGCAAGTGCATCAAGGCTGGCTGTGATGGGAACCTGGGCTCCCAAGAG	2001
Db	2460	GTCTGTGTGCAAGGGCAGTGTGTGAAGCTGGCTGTGATCGCATCATAGACTCCAAAAAG	2519
Qy	2002	AGATTCGACAAGTGTGGGTGTGTGGGGAGACAAATAGAGCTCCAGAGAGTGAAGTGGGA	2061
Db	2520	AAGTTGATGAAGTGTGGCTTTGTGGGAGAAACGGTGTCCATGTCAAGAAAGTGTCAAG	2579

QY	2062	CTCTTCACCAAGCCCATGCAATGCTCAAAATTCGTGGTGGCCATCCCGGCA3GCGCCTCA	2121
Db	2580	ATAGTCACTAGTACAAGACCTGGGTATCATGACATTGTACAAATTCCTGTGGAGCCACC	2639
QY	2122	AGCATCAGACATCCGCCAGCGCGGTTACAAGGGCTGATCGGGGATGACAACACTCTGGCT	2181
Db	2640	AACATTGAAGTGAAACATCCGGAATCAAGGGGGTCCAGAAACAATGGCAGCTTTCTGGCT	2699
QY	2182	CTGAAGAACACGCCAAGCAAGTACCTGCCTCAACGGGCAATTCGTGGTGTGCGGCGTGGAG	2241
Db	2700	ATTAGAGCGCTGATGCTACCTATATTCGTAATGGAACCTTCACCTCTGTCCACACTAGAG	2759
QY	2242	CGGACCTGGTGGTGAAGGCACCTGCTGCGGGTACAGCGGCACGGGCACACGCGTGGAG	2301
Db	2760	CAAGACCTCACCTACAAAGGTACTGCTTTAAGGTACAGTGGTTCCTCGGCTGGCGTGGAA	2819
QY	2302	AGCTGTCAGGCTTCCCGGGCCATCCCTGGAGCGCTGACCGTGGAGGTCTCTCCGTGGGG	2361
Db	2820	AGAAATCGCGAGCTTTAGTCCACCTCAAGAAGCCCTTAACCATCCAGGTTCTTATGCTAGGC	2879
QY	2362	AAGATGACACGCGCCCGGGTCCGCTACTCCTTCTATCTGCCCCAAGAGCCCTCGGAGGAC	2421
Db	2880	CATGCTCTCCGACCCCAAAATTAATTCACCTACTTATGAAGAAGACAGAGATCATTC	2939
QY	2422	AAGTCCTCTC 2431	
Db	2940	AACGCCATTTC 2949	
RESULT 15			
ID	AC90067	standard; DNA; 4180 BP.	
AC	AC90067;		
XX			
DT	19-MAR-2001 (first entry)		
XX			
DE	D67076 cDNA clone.		
XX			
KW	METH; metalloprotease; thrombospondin; angiogenesis inhibition;		
KW	cancer therapy; benign tumour; ocular angiogenic disease;		
KW	rheumatoid arthritis; psoriasis; wound healing; endometriosis;		
KW	vasculogenesis; granulation; hypertrophic scar; nonunion fracture;		
KW	scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;		
KW	coronary collateral; cerebral collateral; arteriovenous malformation;		
KW	ischemic limb angiogenesis; Osler-Webber syndrome; wound granulation;		
KW	plaque neovascularisation; telangiectasia; haemophilic joint; EST;		
KW	angiofibroma; fibromuscular dysplasia; expressed sequence tag;		
KW	Crohn's disease; atherosclerosis; birth control; ss.		
XX			
OS	Unidentified.		
XX			
EN	W0200071577-AL.		
XX			
PD	30-NOV-2000.		
XX			
PF	25-MAY-2000; 2000WO-US14462.		
XX			
PR	25-MAY-1999; 99US-0318208.		
PR	20-JUL-1999; 99US-0144882.		
PR	10-AUG-1999; 99US-0147823.		
PR	13-AUG-1999; 99US-0373658.		
PR	22-DEC-1999; 99US-0171503.		
PR	22-FEB-2000; 2000US-0183792.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(SMK-) SMITHKLINE BEECHAM CORP.		
PA	(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.		
PA	(IRUE/) IRUELA-ARISPE L.		
PA	(HAST/) HASTINGS G A.		
PA	(RUBE/) RUBEN S M.		
PA	(JONA/) JONAK Z L.		
PA	(TRUL/) TRULLI S H.		

PA (FORM/) FORNWALD J A.
 XX (TERR/) TERRETT J A.
 PI Iruela-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Fornwald JA, Terrett JA;
 XX WPI; 2001-025136/03.
 XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis -
 XX Claim 7; Pages 543-546; 768pp; English.
 XX The present invention relates to human METH1 and METH2, (ME for
 CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
 CC The present sequence is an expressed sequence tag (EST) for METH. METH
 CC can be used for inhibiting angiogenesis in an individual, and for
 CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
 CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH can also be used in birth control. METH can also
 CC be used in diagnostic methods for the prognosis of cancer.
 XX Sequence 4180 BP; 1078 A; 1051 C; 1146 G; 905 T; 0 other;
 SQ

Query Match 22.3%; Score 636; DB 22; Length 4180;
 Best Local Similarity 60.7%; Pred. No. 1.3e-122;
 Matches 1098; Conservative 0; Mismatches 700; Indels 12; Gaps 3;

Qy 631 AAGGTTTCGTGTCATCCGCGGTAGCTGGAGACGCTGGTGTGCGGACGAGTCAATG 690
 Db 1143 AAGCCATTTGTGTCACGCCCGCTTATGTGGAACCAATGCTCTAGTGACCACTCAATG 1202
 Qy 691 GTCAAGTTCCACGCGCGGAGTGGAAATATCTGCTGACGCTGTGGCAACGCGGCG 750
 Db 1203 GCCGACTTCCAGGCGAGCGGTCTAAAGCAATACCTTCTAACCCCTGTCTCGTGGCAGCC 1262
 Qy 751 CGACTCTACCGCATCCAGCATCCTCAACCCCATCAACATGCTGTGCTCAAGTGTCTG 810
 Db 1263 AGGTTTACAGCATCCAGCATAGGAATTCATTAATAGCCTGTGTTGGTGAAGATCTTG 1322
 Qy 811 CTTCCTTAGAGATCGTACTCCGCGCCCAAGTCAACGGCAATCGGCGCTGACGCTGCGC 870
 Db 1323 GTCATATACGAGGAGAGAGGACCAAGAGTACCTTCAATGCAGCTCTCACCCCTGG 1382
 Qy 871 AACTTCTGTGCTGGCAGAGAGTGAACAAAGTGAAGTGAACAGCACCCCGAGTACTGG 930
 Db 1383 AATTCTGCACTGGCAGAAACACACACAGCCCAAGTCAACCCGATCCAGAGCACTAT 1442
 Qy 931 GACACTGCCATCCTTTCACGAGGAGGACCTGTGTGGAGCCACCACTGTGACACCCCTG 990
 Db 1443 GACATCGCAATTTGTTTCCAGACAGAGATTTATGTGGCTCCACACAGTGTGACACTCTC 1502
 Qy 991 GGCATGCTGATGGGTGACCATGTGACCCCAAGAGAGCTGCTGTGTCATTGAGGAC 1050
 Db 1503 GGAATGGCAGATGTGAACCGTATGTGACCCCAAGAGAGCTGCTGAGTATAGAGAT 1562
 Qy 1051 GATGGGCTTCCATCAGCTTTCACACTGCCCACAGAGTGGGCCACGTTGTTCAACATGCC 1110
 Db 1563 GATGGTTTGAAGCGCCTTCCACAGCCCATGAATTTGGGCCATGTTTAAACATGCCG 1622
 Qy 1111 CATGACATGTGAAGTCTGTGAGGAGGTGTTGGGAGCTCCGAGAGCCACACATGATG 1170
 Db 1623 CACATGATGTAAGACATGTGGCAGCTTGAATGGTGTGAGTGGCGAATTCATCTATG 1682
 Qy 1171 TCCCGGACCTCATCCAGATGACCGGTGCCAACCCTTGGTGGCTGAGCTGCTGCATC 1230

Db 1683 GCCTCGATGCTCTCCAGCTTAGACCATAGCAGCCCTGGTCACTTGCAGTGCCTACATG 1742
 Qy 1231 ATCACCGACTTCTCTGGACAGCGGCACGGTGACTCCCTCTGGACCAACCCAGCAAGCCC 1290
 Db 1743 GTCAGCTCTTCTTAGATATATGACAGCGGGAATGTTTGTATGGACAAAGCCCAAGATCCA 1802
 Qy 1291 ATCTCCCTGCGGAGGAGTCTGCGGGGCCAGCTACACCTTGAGCCAGCAGTGGAGCTG 1350
 Db 1803 ATCAAGCTCCCTTCTGATCTTCCCGGTACCTTGACATGCAACCCGCAATGTGACGTTT 1862
 Qy 1351 GCTTTTGGTGGGCTCCCAAGCCCTGCTCTTACATGCAATGACAGTAC---TGACCAAGCTGTGG 1407
 Db 1863 ACATTGCGAGAGAAATCCCAAGCACTGCCCTGATGAGCCAGCAGATGACTACTCTGTGG 1922
 Qy 1408 TGCACCGGAAGGCCAAGGAGACAGATGTTGTCAGACCCGCCCACTTCCCTGGGCGGAT 1467
 Db 1923 TGCAGTGGCACTCCGGTGGCTTACTGTGTGTCACCAACAAAACACTTCCCTTGGCAGAT 1982
 Qy 1468 GGCACCACTGTGGCGAGGGAAGCTCTGCTCTCAAGGGGCTGCGTGGAGAGACACAAAC 1527
 Db 1983 GGCACCACTGTGGAGAAAGGAGTGGTGTGTCAGTGGCAAGTGGCTGAACAAAGACAGAC 2042
 Qy 1528 CTCACAAAGCACAG-----GGTGTGATGTTCTCTGGGCAATGGGATCCCTATGGCCCC 1581
 Db 2043 ATGAGCATTTTGTACTCTCTGTTTCATGGAAGCTGGGACCATGGGACCTGGGAGAC 2102
 Qy 1582 TGCTCGCGACATGTGTGGGGGCTGAGCTGCGCAGAGGAGTGCACCAACCCCAACC 1641
 Db 2103 TGCTCAAGAAACCTGTGGTGGAGTTCATATACAAATGAGAGAATGTGACAAACCCAGTC 2162
 Qy 1642 CCGTCCACAGGGGCAAGTACTGCGAGGAGTGGAGGTGAAATACCGATCTGCAACCTG 1701
 Db 2163 CCAAGAGACGGAGGAGTACTGTGAAGCAACAGTCCGCTACAGTCTCTGTAAACATC 2222
 Qy 1702 GAGCCTCTCCCGCAGCTCAGCCTCCGGAAGAGCTTCCGGGAGGAGACAGTGTGAGCTTTTC 1761
 Db 2223 GAGGACTGTCCAGACAATAA---CGGAAAAAGTTTCAGAGAGGAGCAGTGGAGCGCAC 2279
 Qy 1762 AACGGTACAAACACAGCAACCCGCTCACTCTCGCGTGGCATGGTGGCCCAAGTAC-1821
 Db 2280 AATGAGTTTCCAAAGCTTCTCTTGGGAATGAGCCCACTGTAGTGGACACCCCAAGTAC 2339
 Qy 1822 TCCGCGTGTCTCCCGGACAAAGTCAAGCTCATCTCGGAGCAATGGCACTGGCTTAC 1881
 Db 2340 GCGGGTCTCCCAAGAGCAGTGTCAAGCTCACTGTGAAGCCAAAGGCAATGGCTTAC 2399
 Qy 1882 TTCTATGTCTGGCAGCCCAAGGTGTGTGAGCGCAGCTGTGCTCTCTGACTCCACCTCC 1941
 Db 2400 TTTTTCGCTTACAGCCCAAGGTTGTAGTGGCACTCCCTGTGTAGTCCAGACTCTACCTCT 2459
 Qy 1942 GTCTGTGCCAGGCAAGTGCATCAAGCTGGCTGTGTATGGGAACCTGGGCTCCAGAG 2001
 Db 2460 GTCTGTGCAAGGCGAGTGTGTAAGAGCTGGCTGTGTATGCGCATATAGATCCAAAAG 2519
 Qy 2002 AGATTGCAAGTGTGGGTGTGTGGGGAGACAAATAAGAGCTGCAAGAAAGTGAAGTGA 2061
 Db 2520 AAGTTGATAGTGTGGCTTGTGGAGGAACGGTTCACATGCAAGAGATGTCAGGA 2579
 Qy 2062 CTCCTTCAACCAAGCCCATGATGGCTTACAAATTTCTGTGGGCAATCCCGCAGGCGCTCA 2121
 Db 2580 ATAGTCACTAGTACAAAGACCTGGGTATCATGACATTTGTACAAATTCCTGCTGGAGCCACC 2639
 Qy 2122 AGCATGCAATCCCGCAGCGCGGTTACAAAGGCTGTATGGGGATGACAACTACCTGGCT 2181
 Db 2640 AACATTGAAGTGAACATCGGAATCAAGGGGGTCCAGAAAACATGGGAGCTTCTGGCT 2699
 Qy 2182 CTGAAGAACAGCAAGGCAAGTACCTGCTCAACGGGCAATTTCTGGTGTGCGCGGTGGAG 2241
 Db 2700 ATTAGAGCGCTGATGGTACCTATATTCTGATGGAACACTTCACTCTGTCCACACTAGAG 2759
 Qy 2242 CGGGACCTGGTGGTGAAGGCGAGTGTGTGCGGTACAGGGCAGCGGACAGCGGTGGAG 2301
 Db 2760 CAAGACCTCACCTACAAAGGTACTGTCTTAAGGTACAGTGGTTCCTCGGCTGCGCTGGAA 2819

QY 2302 AGCTGAGGCTTCCCGCCCATCTGGAGCGCTGACCGTGGAGGTCTCTCCGTGGG 2361
 Db 2820 AGAATCCGACGTTTGTCCACTCAAGAACCCTTAACCATCCAGGTTCTTATGGTAGGC 2879
 QY 2362 AAGATGACACCGCCCGGTCCTCTTCTATCTGCCCAAAGAGCCTCGGGAGGAC 2421
 Db 2880 CATGCTCTCCGACCCAAAATTAATTCACCTACTTTATGAAGAAGACAGAGTCATTC 2939
 QY 2422 AAGTCCTCTC 2431
 Db 2940 AACGCCATTC 2949

Search completed: May 1, 2003, 19:41:34
 Job time : 639 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 19:29:16 ; Search time 103 seconds
(without alignments)
8494.649 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgctttgctgggcatcct.....gcgtcttgaggcgtgctga 2853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	628.2	22.0	4676	4	US-09-130-491-1
2	618.8	21.7	4858	4	US-09-392-184-1
3	613.3	21.5	3708	4	US-09-484-970B-58
4	599.8	21.0	3126	4	US-09-392-184-7
5	592	20.8	3638	4	US-09-369-364A-8
6	584.8	20.5	4192	4	US-09-122-126B-1
7	547.4	19.2	3250	4	US-09-122-126B-14
8	545.8	19.1	3002	4	US-09-369-364A-1
9	471.4	16.5	2825	4	US-09-369-364A-14
10	411.2	14.4	5804	4	US-09-369-364A-12
11	318	11.1	2114	4	US-09-130-491-7
12	259.4	9.1	739	4	US-09-369-364A-10
13	247.2	8.7	5357	4	US-09-392-184-5
14	240.2	8.4	3218	4	US-09-369-364A-6
15	215.8	7.6	1520	4	US-09-369-364A-3
16	214.4	7.5	3885	4	US-09-369-364A-16
17	187	6.6	2450	4	US-09-491-522-2
18	187	6.6	6692	4	US-09-491-522-1
19	179.4	6.3	2450	4	US-09-491-522-9
20	179.4	6.3	4580	4	US-09-491-522-8
21	162.2	5.7	703	4	US-09-392-184-6
22	117.6	4.1	3675	4	US-09-930-872-3
23	117.6	4.1	4042	4	US-09-930-872-5
24	91.6	3.2	2848	4	US-09-369-364A-4
25	79	2.8	2023	4	US-09-491-522-6
26	77	2.7	385	4	US-09-392-184-23
27	67.6	2.4	349	4	US-09-392-184-24

28	61.4	2.2	1642	4	US-09-369-364A-18	Sequence 18, Appl
29	57	2.0	1476	4	US-09-930-872-1	Sequence 1, Appli
30	55.8	2.0	1803	4	US-09-369-364A-20	Sequence 20, Appl
31	50.4	1.8	657	3	US-08-985-526-2	Sequence 2, Appli
32	50.4	1.8	1326	3	US-08-985-526-4	Sequence 4, Appli
33	49.4	1.7	716	4	US-09-276-531-117	Sequence 117, App
34	47.4	1.7	3014	2	US-08-808-982-1	Sequence 1, Appli
35	47.4	1.7	3014	2	US-09-306-902A-1	Sequence 1, Appli
36	47.2	1.7	1173	3	US-08-706-216-5	Sequence 5, Appli
37	46.4	1.6	1101	4	US-09-152-060-23	Sequence 23, Appl
38	44.6	1.6	502	4	US-09-392-184-12	Sequence 12, Appl
39	44	1.5	7452	3	US-08-592-500-1	Sequence 1, Appli
40	44	1.5	7452	3	US-08-195-006-1	Sequence 1, Appli
41	44	1.5	7452	5	PCT-US94-07644A-1	Sequence 1, Appli
42	43.6	1.5	2063	1	US-08-715-554-1	Sequence 1, Appli
43	43.6	1.5	2063	2	US-08-583-118-1	Sequence 1, Appli
44	43.4	1.5	7218	1	US-08-232-463-14	Sequence 14, Appl
45	43	1.5	20235	1	US-07-642-734C-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-130-491-1
; Sequence 1, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4676
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (460)....(3360)
US-09-130-491-1

Query Match	22.0%	Score 628.2;	DB 4;	Length 4676;
Best Local Similarity	57.1%	Pred. No. 1.1e-127;		
Matches 1305;	Conservative 0;	Mismatches 913;	Indels 69;	Gaps 6;
QY	175	GCATTTGAGGAGGACTTTTACCTACACCTGAGCGGGATGCTCAGTCTTGGCTCCGCC	234	
DB	697	GCCTTTGACGAGCTGGATCTGGAGCTGGCGCCGACAGCAGCTTTTGGCCCGCCG	756	
QY	235	TTCTCACTGAGCACTCTGGGGTCCCTCCAGGGGCTCAGCGGGGCTC---TTCAGAC	291	
DB	757	TTACAGCTCCAGAGCTGGGGCGCAATCCGGGTCGAGAGACCGCTTCCGGAACCGAC	816	
QY	292	CTGGAGCCTGCTTCTATTCTTGGGAGCTGAACCGCGGAGCTCGTTGCTGCTGTG	351	
DB	817	CTGGCGCACTGCTTCTACTCCGGCACCGTGAATGCGATCCAGTCCGCTCGCGCCCTC	876	
QY	352	AGCCTGTGCGGGGGCTCCGGGAGCCTTTGGCTACCGAGGGCGGAGTATGCTATAGC	411	
DB	877	AGCCTCTCGAGGGGCTCGCGCGCCCTTACCTGCTGGGGGAGGCGTATTTTCATCCAG	936	
QY	412	CGCTGCCCAATGCTAGCGCGC-CGGCGGCGCAGCGCAACAGCCAGGGCGCACACCTTCT	470	
DB	937	CGCTGCCCGCGGCGGAGCGGCTCTGCCACCGCCGCCCGCCCGGAGAACCCCGGCA	996	

QY 471 CAGAGCCGGGGTGTTCGGGGGGGGCTTCGAGAGACCCCACTCTCGTTCGGGGGGTGGC 530
DB 997 CCACTACAGTTCCACCTCTCGGGGGGAAATAGCAGGGGAGCTAGCGGCACAGTGGCGG 1056
QY 531 CTCGGGCTGGAACCCCGCCATCTAGGGCCCTTGACCCCTTACAAGCCGCGCGGGGG 590
DB 1057 GTCTGTGACGAGAGCCCGGGGCTGAGGAAAGGGAGAGCGAAGAGAGGAGGAAAGG 1116
QY 591 CTTTCGGGGAGAGTCGTAGCGGGCGAGGTCTGGGGC----- 627
DB 1117 ACTGAGGGGAGGACGAAGGGCTCAGTGTGCGCCGAGGACCCGGCACTGCAAGCGTA 1176
QY 628 -----GCCAAGCGTTTCGTGTCTATCCCGGGTAC 657
DB 1177 GGAGAGCCACAGGAACCTGGAAGCATAAGAAAGAGCGATTGTGTCCAGTCAACGGCTAT 1236
QY 658 GTGGAGACGCTGGTGTGCGGAGCAGTCAATGGTCAAGTTCACGGGCGGACCTGGNA 717
DB 1237 GTGGAACCATGCTTGTGGCAGACCAAGTGTGGGAGAAATCCACGGGAGTGGTCTAAG 1296
QY 718 CATTATCTGCTGACGCTGTGGCAAGCGGGCGGCACTCTACCGGCATCCCGAGCATCTC 777
DB 1297 CATTACCTTCTCAGCTGTGTTTCGGTGGCAGCCAGATTGTACAAACACCCAGCATTCGT 1356
QY 778 AACCCCATCAACATCTTGTGTCAAGGTGCTGCTTTAGAGATGTGACTCCGGGGCC 837
DB 1357 AATTCAGTTAGCTGTGGTGTGAAGATCTTGGTCAATCCAGATGAACAGAAAGGGCGC 1416
QY 838 AAGTCCACCGGAATCGCGCCCTGAGCTGCGCAACTCTGTGCTCGCAGCAAGAGCTG 897
DB 1417 GAAGTCACTTCAATGCTGCCCTCACTGTGGGAATTTTGAACCTGGCAGAGCAGCAC 1476
QY 898 AACAAAGTGTGACAAAGCACCCGAGTACTGGGCACTGCGCATCTCTTACCAGGAGC 957
DB 1477 AACCCACCCAGTGACCGGATGCAGAGCACTATGACAGCAATTCCTTTCACAGACAG 1536
QY 958 GACCTGTGTGGAGCCACCACTGTGACACCCCTGGGATGGCTGTGATGGGTACCATGTGT 1017
DB 1537 GACTGTGTGGTGTCCAGACATGTGATCTTGGGATGGCTGTGATGTTGGAATCTGTGT 1596
QY 1018 GACCCCAAGAGAGTGTCTGTCTATTGAGAGCATGGCTTCCATCAGCCTTCAACCACT 1077
DB 1597 GATCCAGCAGAGTGTCTCCCTCATAGAAGATGATGTTTACAAGCTGCTTCAACACA 1656
QY 1078 GCCCAGAGCTGGGCACTGTTCAACATGCCCACTGACAAATGTGAAAGTCTGTGAGGAG 1137
DB 1657 GCCCATGAATTAGGCCACGCTTTAATCATGCCACATGATGATGACAAAGCAGTGTGCCAGC 1716
QY 1138 GTGTTTGGAGCTCGGAGCAACACATGATGTCGCCGACCTCATCCAGATCCAGCT 1197
DB 1717 CTTAATGGTGTGAACCCAGGATTTCCACATGATGGCTCAATGCTTCCAACTGGACCAC 1776
QY 1198 GCCAACCCCTGTGTCAGCTGCAAGTGTGCAATCATCAGCACTTCTCGACAGCGGGCAQ 1257
DB 1777 AGCCAGCTTGTCTCTTGTGAGTGCCTACATGATTTATCATCTTCTGGATAATGGTCTAT 1836
QY 1258 GTGACTGTCTCTGGACAAACCCAGCAAGCCATCTCCCTGCCGAGGATGTGCCGGGC 1317
DB 1837 GGGGAATGTTGATGGACAAGCCTCAGAAATCCATACAGTCCCAAGGCGATCTCCCTGGC 1896
QY 1318 GCCAGTACACCTGAGCCAGCAGTGGAGCTGGCTTTTGGGCTGGGCTCCAGGCCCTGT 1377
DB 1897 ACTCGTAGGATGCCAACCGGAGGTGCCAGTTTACATTTGGGGAGGAGCTCCAAACACATGC 1956
QY 1378 CTTTACATCAGTAC---TGCACCAAGTGTGGTGCACCGGGAAGGCCAAGGAGACATG 1434
DB 1957 CCGGATGACGAGCAGACATGTAGACCTTGTGTGTACCGGACCTCTGTGGGGTGTG 2016
QY 1435 GTGTGCCAGACCCGACATTCCTCCCTGGGCGGATGGCACCAGCTGTGGGAGGCAAGCTC 1494
DB 2017 GTGTGTCAAAACCAACACTTCCCGTGGCGGATGGCACCAGCTGTGGAGAAGGAAATGG 2076
QY 1495 TGCCCTAAAGGGCCCTGGCTGGAGAGACACACACCTCAACAAG-----CACAGGTGGAT 1548

DB 2077 TGTATCAACGCGCAAGTGTGTGAACAAAACCCAGACAAGCAATTTTGATACGCCCTTTTCAT 2136
QY 1549 GGTTCCTGGGCAAAATGGGATCCCTATATGGCCCTGTCTCGCGCACATGTGTGGGGGGTG 1608
DB 2137 GGAAGCTGGGAATGTGGGGCTTGGGAGAGATGTTCGAGAAAGCTGGGTGGAGGATC 2196
QY 1609 CAGTGGCCAGGAGGAGTGTGACCAACCCCACTCCCAACGGGGCAAGTACTGGGAG 1668
DB 2197 CAGTACAGATGAGGGAATGTGACAAACCCAGTCCCAAGAAATGGAGGAAAGTACTGTGAA 2256
QY 1669 GGAGTGGAGGTGAATACCGATCTGCAACTGGAGCCCTGCCCCAGCTCAGCCTCCGGA 1728
DB 2257 GGCAAGAGAGTGCCTACAGATCTGTAACTTGAGGACTGTCCAGAC---AATAATGGA 2313
QY 1729 AAGAGCTTCGGGAGGAGCAGTGTGAGCTTCAACCGCTTACAACGACACACCAACCGG 1788
DB 2314 AAAACCTTTAGAGAGGAAATGTGAAGCACACAAAGGATTTTCAAAAGCTTCTCTTTGGG 2373
QY 1789 CTCACTCTCGCCGTGGATGGGTGCCCAAGTACTTCGGGCGTGTCTCCCGGGACAAAGTGC 1848
DB 2374 AGTGGCTCGGGTGAATGATTTCCCAAGTACGCTGGCTCTCAACCAAGGACAGGTGC 2433
QY 1849 AAGCTCATCTGCCGAGGCAATGGCACTGGCTACTTCTATGTGTGGCACCCCAAGTGGTG 1908
DB 2434 AAGCTCATCTGCCAAGCAAGGCAATGGCTACTTCTTCTGCTTTTGCAGCCCAAGGTGTA 2493
QY 1909 GACGGCAGCTGTCTCTCTGACTCCACCTCGCTGTGTCGCAAGGCAAGTGCATCAAG 1968
DB 2494 GATGGTACTCCTATGCTAGCCAGATTTCCACCTCTGCTGTGTCAGAGCAGTGTGTA 2553
QY 1969 GCTGGCTGTGATGGGAACCTGGGCTCCAAGAAGAGATTCGACAAAGTGTGGGTGTGGG 2028
DB 2554 GCTGTTGTGATGCTATGATAGACTTCCAAAAGAAAGTTTGATAAATGTGTGT---TGGGG 2613
QY 2029 GGACACATATAGAGCTGCAAGAAGTCACTGGACTCTTACCAAGCCCATGCAATGCTAC 2088
DB 2614 GGAAATGGATCTACTTGTAAAAAATATCAGGATCAGTTACTAGTGAACCACTGGATAT 2673
QY 2089 AATTTCTGTGTGGCATCCCGGAGCGCTCAAGCATCGACATCCGCCAGCGGTTAC 2148
DB 2674 CATGATATCATCAATTTCCAACTGGAGCCCAACACATCGAAGTGAACAGCGGAACAG 2733
QY 2149 AAAGGCTGTATCGGGATGACACTACTGGCTGTGAAGAACAGCAAGCAAGTACCTG 2208
DB 2734 AGGGATCCAGGAACAATGGCAGCTTTCTGCCATCAAGCTGTGATGSCATATATT 2793
QY 2209 CTCAACGGCATTTCTGTGTGTGGGCTGAGGGAGCCTGGTGGTGAAGGCGAGTCTG 2268
DB 2794 CTTAATGGTACTACATTTGTCCACCTTAGAGCAACATATGTACAAGGTGTGTGTC 2853
QY 2269 CTGGGTACAGCGGACGCGCACAGCGGTGGAGCGCTTCAGGCTTCCCGGCCCATCCTG 2328
DB 2854 TTGAGGTACAGCGCTCTCTCGGCATTTGAAAGATTCGCAAGTTCAGCCCTCTCAA 2913
QY 2329 GAGCGCTGACCTGGAGGTGCTCTCGTGGGGAAGATGACACCGCCCGGGTCCGCTAC 2388
DB 2914 GAGCCCTTGACCATCCAGTCTTACTGTGGCAATGCCCTTGCAGCTAAATTAATATAC 2973
QY 2389 TCCTTCT 2395
DB 2974 ACCTACT 2980

RESULT 2

US-09-392-184-1

; Sequence 1, Application US/09392184

; Patent No. 6395889

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN

; TITLE OF INVENTION: PROTEASE HOMOLOGS

; FILE REFERENCE: 5800-55


```

Db 1954 CCGGATTTCCGCGACATGCCCCAACACCTCTGTCTAGGAGCTCTGGCCCCAGCTTTGGTGC 1895
QY 1408 TGCACCGGAAGCCAAAGGACACAGATGGGTGTCAGACACCGCCACTTCCCTGGGCGCAT 1467
Db 1894 CACACTGATGGGCTGAGCCCTGTGCCACAGAAATGGCAGCTGCCCTGGGCTGAC 1835
QY 1468 GGCACCAAGCTGTGGCGAGGCGAAGCTCTGCTCTCAAAAGGCGCTGC-----GTGGAGAG 1520
Db 1834 GGCACCGCTGTGGCGCTGGCCACCTCTGCTCAGAAGGCAAGCTGTCTACCTAGGAGGA 1775
QY 1521 ACACAACTCAACAGCAGAGGCTGATGGTTCCTGGGCGCAATGGGATCCCTATGGCCC 1580
Db 1774 AGTGGAGAGCCCAAGCCCTGTAGATGAGGCTGGGACCGCTGGGACCTGGGGAGA 1715
QY 1581 CTGCTCGGCACATGTGGTGGGCGCTGCAGCTGGCCAGGAGCAGTGCACCAACCCAC 1640
Db 1714 ATGTTCTCGGACCTGTGGAGGAGAGTACAGTTTTCACACCGTGTAGTGAAGACCCGA 1655
QY 1641 CCTGCGCAACGGGGCAAGTACTGCGAGGAGTGGGTTGAATACCGATCTCTGCAACCT 1700
Db 1654 GCCTCAGAATGGAGGAATACCTGCTGGTGGGAGGAGCAAGTACCACTCATGCCACAC 1595
QY 1701 GGAGCCCTGCCAGCTCAGCCTCGGAAAGAGCTTCCGGGAGGACAGTGTGAGCTTT 1760
Db 1594 GGAGGAATGCCCC-----CCTGACGGGAAAGCTTCAGGAGCAGCAGTGTGAGAGTA 1541
QY 1761 CAACGGCTACACACACAGCACCACCGGCTCACTCTGCCGCTGGCATGGGTGCCCAAGTA 1820
Db 1540 TAATGCTCAATATACACTGACATGACATGGAGGGAATCT---CCTCAGTGGTCCCAAGTA 1484
QY 1821 CTCGGGCGTCTCCCGGACAGATGCAAGCTCATCTGCCGAGCAATGGCACTGGCTA 1880
Db 1483 TGCTGGGGTGTCCCGCGGACCGCTGCAAGTTGTTCTGCCGAGCCCGGGGAGGAGCGA 1424
QY 1881 CTTCTATGTGTGCGACACCAAGTGTGGAGCGCAGCTGTCTCTCTGACTCCACCTC 1940
Db 1423 GTTCAAGTGTTCGAGCCCAAGTGATGATGGCACCTGTGTGGGCCAGAACACTGCG 1364
QY 1941 CGTGTGTCCAAAGCAAGTGCATCAAGGCTGGCTGTGATGGGAACCTGGGCTCCAGAA 2000
Db 1363 CATCTGTGCTGGCCAGTGTGTCAAGGCGGCTGTGTGACCATGTGTGGACTCGCTCG 1304
QY 2001 GAGATTCGACAGTGTGGGTGTGTGGGGAGACAAATAGAGCTGCAAGAGTGTGACTGG 2060
Db 1303 GAAGCTGGACAAATGCGGGGTGTGTGGGGCAAGGCAACTCTGCAGGAAGTCTCCGG 1244
QY 2061 ACTCTTCAACCAAGCCCATGATGCTACAATTTGCTGGTGGCCATCCCGCAGCGCCCTC 2120
Db 1243 GTCCCTCACCCCCACCAATATGGCTACATGACATTTGTCACCATCCACGCTGTGGCAG 1184
QY 2121 AAGCATCGACATCCCGCAGCGGTTTACAAGGCTGATCGGGGATGACAACTACCTGGC 2180
Db 1183 TAATATTGACGTGAAGCAGCGGAGCCACCGGGTGTGTGACAAAGTGTGGAACTACCTGGC 1124
QY 2181 TCTGAAGACACCAAGCAGTACCTGCTCAAGGGCAATTTGCTGTGTGCGGGTGA 2240
Db 1123 GCTGAAGACGCTGATGGGAGTACCTGCTCAAGGGCAACTTGGCCATCTCTGCCATAGA 1064
QY 2241 CCGGAGCTGTGTGAGGAGCAGTCTGCTGCGGTACAGGGGACAGCGGACAGCGTGA 2300
Db 1063 CGAGGACATCTTGTGAAGGGAGCACCCTCTGAAGTACAGCGGCTCCATCGCCACCTGGA 1004
QY 2301 GAGCCTGAGGCTTCCCGGCGCCATCTCTGGAGCGCTGACCGTGGAGTCTCTCTCCGT--- 2357
Db 1003 CGCGCTGAGAGCTTCCGGCGCTTGGCAGAGCCTTGACAGCTGAGCTCTCTGACAGTCCC 944
QY 2358 GGGGAGATGACACGGCCCGGGTCCGCTACTCTCTCTATCTGCGCCAAAGA 2408
Db 943 TGGGAGGTCTTCCCGCCCAAAAGTCAATACACCTTTCTTGTCTTATGA 893

```

RESULT 5
US-09-369-364A-8

```

; Sequence 8, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 3638
; TYPE: DNA
; ORGANISM: Mus musculus ADAMTS-8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (278)..(2992)
; NAME/KEY: misc_feature
; LOCATION: (3636)
; OTHER INFORMATION: n = T
US-09-369-364A-8

```

```

Query Match 20.8%; Score 592; DB 4; Length 3638;
Best Local Similarity 57.3%; Pred. No. 7.8e-120; Indels 90; Gaps 11;
Matches 1335; Conservative 0; Mismatches 905;
QY 157 CTCATTTTTCAGATCACAGCATTTTCAGGAGACTTTTACCTACACCTGACCCCGGATGCT 216
Db 440 CTCGCTTCCACCTCTCCGCTTCGCCAGGCTTTCGTCTCGCTGGCCCTGAGGCC 499
QY 217 CAGTTCTTGGCTCCCGCTTCTCCACTGAGCATCTGGCGTCCCTCCAGGGCTACC 276
Db 500 AGCTTCTTGGCGCGGAATTAAGATCGAGCGCTTCGGGGCTCGAGCGCGCG---GCC 556
QY 277 GGGGGCTTTCAGACTCGGACGTCTTCTTATCTTGGGGAGCTGAACCGCGAGCGGAC 336
Db 557 GGGGGGAGCGGGACTCGTGGCTGCTTCTTCTGACAGTGAATGGAGACGGGAG 616
QY 337 TCGTTCGTGTGTGAGCCTGTGGGGGGGCTTCGCGGAGCCTTTGGCTACCGAGCGCC 396
Db 617 TCGCTGGCGGATGAGTGTCTCGCGGCTGGAGCGCTCTGCTTCTTGGCAGCGGAG 676
QY 397 GAGTATGTCATTAGCCCGCTGCCCAATGCTAGCGCGCGCGGCGCAGCGCAACAGCCAG 456
Db 677 GAGTTCACCATTCAGCACAGCGCGCTGGGGACTCCCTGGACCACCTTCATCGCTGAG 736
QY 457 GCGGCACACCTTCTCCAGCGCGGGGTGTTCGCGGCGGGGCTTCCGGAGA---CCCCAC 512
Db 737 CGCTGGGGCGCGGACAGCGCGCGAAGACCCCGGGCTCGCTGCGCGCGAAGTTTTCGCC 796
QY 513 CTCCTGCTCGGGGTGGCCTCGGGGTGGAAACCCCGCATCTTACGGGCCCTGGACCTTA 572
Db 797 CTCCTTCAAGGACTGAGTGGAGGTGGAGATGGTAAATGGCGAGGACAGGAGAAAGT 856
QY 573 CAGCGCGCGCGGGGCTTCGGGAGAGTGTAGCGCGCGCAGTCTGGG-----G24
Db 857 GACAACAAAGAGACAAAGAACAGGAGGGGTGCTCAAAGAGACAGAACTCC 916
QY 625 -----CGCGCCAAAGCGTTTGTGTCCTATCCCG 651
Db 917 CGCAAGTGCACACCTTCGGATCCAAATAGAAAGCAAGAGTTTGTGTCGAGGCT 976
QY 652 CGGTAGCTGGAGACGCTGTGGTGGCGGAGAGTCAATGGTCAAGTCCACGCGCGGAG 711
Db 977 CGCTTCTGGAACACTTCTGTGGTGTGATGCTGCTCATGGCTGCTTATGGGCCGAC 1036
QY 712 CTGGAACATTATCTGCTGAGCCTGTGCAACGGCGGCGGCTCTACCGCATCCAGC 771
Db 1037 CTGCAACCAATCTCTCAGGTGATGTCATGGGAGCCGATCTACAAGCACCGGAG 1096
QY 772 ATCCTCAACCCCATCAACATCGTTGTGGTCAAGGTGCTGCTTCTTAGAGATCGTACTCC 831

```

```
Db 1097 ATCAGGAACCTCCGTCACAACTTGTGGTGAAGTCTAATAGTGAAAGAAAGATGG 1156
QY 832 GGGCCCAAGTTCACCGCAATGGGCGCCCTGACGCTCGGCAACTTCCTGTCCTGGGAGAAG 891
Db 1157 GGGCCCGGAAGTGTCCGACAAACGGGGGGCTCACACTCGCAACTTCTGCACTGGCAACGG 1216
QY 892 AAGCTGAACAAGTGAAGTACAGAACCCCGAGTACTGGAGCACTGCCATCTCTTCACC 951
Db 1217 CGTTTCAACAAGCCAGTACCCGCAACCCGGAGCACTATGACACTGCCATCTTGTTCACC 1276
QY 952 AGCAGAACCTGTGTGG---AGCCACCACTGTGACACCTCGGGCATGGCTGATGTGGGT 1008
Db 1277 AGACAGAACAATCTGTGGGAAGGAGAGACAGTGTGACACCCCTGGGATGSCAGAGTGGC 1336
QY 1009 ACCATGTGTACCCCAAGAGAGAGTCTCTGTCTGATGAGAGAGATGGGCTTCATCAGCC 1068
Db 1337 ACCATGTGTACCCCAAGAGAGTCTCTGATGATCAAGGATGAGGAGTGCAGGACGCC 1396
QY 1069 TTCACCACTGCCACAGAGCTGGGCCACGTTTCAACATGCCCATGACAAATGTGAAGTC 1128
Db 1397 TACACCTGGCCCATGAGCTAGGGACGTTCTCAGCATGCCCATGATGATTTCTAAGCCC 1456
QY 1129 TGTGAGAGAGTGTGGGAAGTCTCGAGCCCAACACATGATTCGCCCGACCCCTCATCCAG 1188
Db 1457 TGTGTGAGATTTGTGGGCCCATGGCAAGTATCCACATGATGGCGCCATTTCTTCATCCAC 1516
QY 1189 ATCGACCGTGCACAAACCCCTGTCAGCTGCAGTCTGCTGATGACATCATCAGGACATTCCTGAC 1248
Db 1517 GTGAACAAGACGCTGCCCTGGTCTCCCTGAGTGTGTCTACCTCAGAGTCTCTGGAT 1576
QY 1249 AGCGGCACGGTGACTGCTCTCTGGACCAACCCAGCAACCCCATCTCCCTGCGCCGAGGAT 1308
Db 1577 GATGGTCAACGAGATTGTCTCTGGATGCCCACTCGGTTCTGCCCTCCCAAGCC 1636
QY 1309 CTGCGGGGCGCCAG-----CTACACCTTGAGCCAGACAGTGGAGTGTCTTTTGGCGTG 1362
Db 1637 CTCGCGGGGCACAGCACCCCTCTACGAGCTGGACAGCAGTGAAGCAGATCTTTTGGGCT 1696
QY 1363 GGCTCCAAGCCCTGTCTTTACA-TGCAGTACTGCACCAAGCTGTGTGCACCGGAAGGC 1421
Db 1697 GATTTCCGACACTGCCCAACACCTCTGTGGAGGACATCTGTCCAGCTCTGTGCCGT 1756
QY 1422 CAAGGACAGAGTGTGTGCC-----AGACCCGCCACTTCCCTTGGCCGAT 1467
Db 1757 CATCGGGATAGTATGAGCCCATTTGCCACAAAGATGTGAGCTGTCTTGGCTGTAT 1816
QY 1468 GGCACAGCTGTGGGAGGCAAGCTCTGCCTCAAGGGGCTGTGTGGAGAGACACAA 1527
Db 1817 GGTACACCTGTGGCCCTGGGACCTGTGCCTGGATGTGTGTACTCAAGGAGGAT 1876
QY 1528 CTCACAAAGCACA-----GGTGTGATGTTCTTGGGCCAAATGGATCCCTATGGCCCC 1581
Db 1877 GTGGAAATCCCAAGCTGTGTGTAGATGAGACTGGGTCTCTGGAGACCTTGGGGACA 1936
QY 1582 TGTCCGCAATGTGTGGGGGTGTGAGCTGTGGCCAGAGGACAGTGCACCAACCCCAACC 1641
Db 1937 TGTTCCTCGACCTGTGTGGAGGATACAAATCTCGAACCCGTGAATGTGATAATCCAATG 1996
QY 1642 CTGCCCCAGGGGCAAGTACTGCGAGGAGTGAAGTGAATACCGATCTCTGCAACCTG 1701
Db 1997 CTTCAAGTGGAGAGATTTTGGCTGGTGAAGAGTCAAGTACCAATCATGCAACACA 2056
QY 1702 GAGCCCTGCCAGCTCAGCCCTCCGAAAGAGCTTCCGGGAGGACAGTGTGAGGCTTTC 1761
Db 2057 GAGGAATGTC-----CACCACGAAAGAAAGCTTCCGGGACACAGTGTGAGAAATAT 2110
QY 1762 AAGGGTACACCAAGACACCAACACCGGTCTACTCTGCGGTGGATGGGTGCCCAAGTAC 1821
Db 2111 AATGCTTACAACCA-----CACTGACCTGGATGGAAATTTCTTGCAGTGGGTCCCCAAGTAT 2167
QY 1822 TCGGGGTGTCTCCCGGGACAAAGTGCAGCTCATCTGCCGAGCCAAATGGGCACTGGCTAC 1881
```

```
Db 2168 TCAGGAGTGTCCCCCGAGACCGATGCAAGCTGTTTTCAGAGCCCGTGGAGAGTGA 2227
QY 1882 TTCATATGTGTGGCACCAAGTGTGTGGACGCGACGCTGTCTCTCTGTGATTCACCTCC 1941
Db 2228 TTCAAAGTGTTTGAAGCTTAAGTGTATCGATGGCACTCTGTGTGGACGGATACTCTGTCC 2287
QY 1942 GTCTGTGTCCAAAGCAAGTGCATCAAGCTGGCTGTGTGATGGGAACCTGGGCTCCAAAG 2001
Db 2288 ATCTGCTGCCGGGGCAATGTTAAGCTGGCTGTGACCATGTGGTGAATCACTACCTAAG 2347
QY 2002 AGATTTCGAAAGTGTGGGGTGTGTGGGGAGACAATAAGAGCTGCAAGAAGCTGACTGGA 2061
Db 2348 AAGCTGGACAAATGTGGGGTGTGTGGGGCAAGGCACTGCTCTAGGAAGATCTCCGGT 2407
QY 2062 CTCCTTCAACCAAGCCATGCATGGCTACAAATTTCTGTGTGGCCATCCCGGACGGCCCTCA 2121
Db 2408 TCTTTTACCCCTTTCAGTTATGGCTTACAAATGACATTTGTCACTCCAGCTGGTGGCACA 2467
QY 2122 AGCATCGACATCCCGACGCGGTTTACAAAGGCTGTATCGGGGATGACAACTACCTGGCT 2181
Db 2468 AACATGTATGTGAACACGCGAGTCAACAGGGTTCAGGAACGACGACGCTACTCTGGCG 2527
QY 2182 CTGAAGAACAGCCAAAGCAAGTACTGCTCAACGGGCAATTTCTGTGTGGCGGTGGAG 2241
Db 2528 CTGAAGACAGCAATGGGCACTACTGCTCAATGTAACTGGCCATCTCTTGCCATAGAG 2587
QY 2242 CGGACCTGGTGTCAAGGGCAGTCTGCTGGGTTACAGCGGCACGGGCACAGCGGTGGAG 2301
Db 2588 CAAGACATCTTGGTGAAGGGGACCATCTTGAAGTACAGTGGCTCCATGGCTTACCCTGGAG 2647
QY 2302 AGCTGTGAGGCTTCCCGGCCCATCTGTGAGCCGCTGACCGTGGAGGTCTCTTCCGTG--- 2358
Db 2648 CGGCTGCAGAGCTTCCAGGCCCTGCTGAGGCTCTTACAGTACAGCTCTCTGACTGTCT 2707
QY 2359 GGGAGATGACACCGCCCGGTCGGTACTCTTCTTATCTGCCCCAAGA 2408
Db 2708 GGTGAGGTCTTCCCTCCAAAAGTCAGATATACCTTCTTTTGTCCCAATGA 2757
```

RESULT 6

US-09-122-126B-1

; Sequence 1, Application US/09122126B

; Patent No. 6451575

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: AGGRECAN DEGRADING METALLO PROTEASES

; FILE REFERENCE: DM6909

; CURRENT APPLICATION NUMBER: US/09/122.126B

; CURRENT FILING DATE: 1998-07-24

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 4192

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (406)..(2916)

US-09-122-126B-1

Query Match

Best Local Similarity 20.5%; Score 584.8; DB 4; Length 4192;

Matches 1094; Conservativity 60.4%; Pred. No. 3e-118;

Mismatches 692; Indels 24; Gaps 7;

```
QY 609 CCGCGCAGGTCTGGGCGGCCAAGCGTTTCTGTCTATCCCGGTTACGTGGAGAGCT 668
Db 1014 CAGCCCCAGACCCCGAAGAGCAAGCGCTTGTCTTACTAGTAGATTTTGTGGAGACT 1073
QY 669 GGTGCTCGCGACAGTCAATGGTCAAGTTTCCACGGCGGACCTGGAACTATCTGCT 728
Db 1074 GGTGCTGGCAGATGCAAGATGGCCGATTCACGGTGGGGGCTAAAGGCTTACTGCT 1133
QY 729 GACGCTGCTGGCAACGGCGCGGACTCTACCGCATCCAGCATCTCAACCCCAATCAA 788
```

Db	1134	AAACAGTATGGCAGCAGCAAGGCCCTTCAAGCACCCCAAGCATCCGAATCCTGTCCAG	1193
Qy	789	CATCGTTGTGGTCAAGGTGCTTCTTAGAGATGTGACTCCGGGGCCCAAGTGCACCG	848
Db	1194	CTTGGTGGTACTCGGCTAGTATGCTTGGGTGAGCGAGGAGGGGCCCAAGTGGGGCC	1253
Qy	849	CAATGGGGCCCTGACGCTGGCGAACTTCTGTGCTGGCAGAGAAGCTGAACAAAGTGAG	908
Db	1254	CAGTGTGCCACAGACCCCTCGCAGCTTCTGTGCTGGCAGCGGGGCCCTCAACACCCCTGA	1313
Qy	909	TGACAGCACCCCGAGTACTGGGACACTGCCATCCTCTTCAACAGCAGGACACTGTGTG	968
Db	1314	GGACTCGGACCCTGACCACCTTTGACACAGCCATTTCTGTTACCCGTCAGGACCTGTGTG	1373
Qy	969	AGCCACACCTGTGACACCCCTGGGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAG	1028
Db	1374	AGTCTCCACTTGCACACCGCTGGGTATGSGCTGATGTGGCACGCTGTGACCCGCTCG	1433
Qy	1029	AAGCTGCTGTGCATTTGAGGACGATGGCTTCCATCAGCCTTCACTACTGCCACAGCT	1088
Db	1434	GAGCTGTGCCATTTGTGGAGATGATGGCTCCAGCTCAGCCTTCACTGCTCATGAAC	1493
Qy	1089	GGCCACAGTGTTCACATGCCCCCATGACAATGTGAAAGTCTGTGAGGAGTGTTTGGGAA	1148
Db	1494	GGGTGATGCTTCAACATGCTCCATGACAACTCCAAGCCATGCATCAGTTTGAATGGCC	1553
Qy	1149	GCTCCAGGCCAA---CCACATGATGTCCCGACCCCTCATCCAGATCCAGCTGCCAACCC	1205
Db	1554	TTTGACACACTCTCGCCATGTCAATGGCCCTTGTATGGCTCATGTGATCCTCAGGAGCC	1613
Qy	1206	CTGGTCAGCTGCAGTGTGCCATCATCAGCACTTCTCTGGACAGGGCACGCTGACTG	1265
Db	1614	CTGGTCCCTCGAGTGGCCGCTTCACTACCTGACTTCTTGGACAAATGGCTATGGGCACTG	1673
Qy	1266	CTCTCTGGACCAACCCAGCAGCCCATCTCCCTCCCGGAGATCTGCCGGGCCACAGCTA	1325
Db	1674	TCCTTTAGACAACACAGAGGCTCCATTCATCTGCCTGTGACTTTCCTTGGCAAGACTA	1733
Qy	1326	CACCTTGACCGACGAGTGGAGCTGGCTTTTGGGTGGGTCCCAAGCCCTGTCTTACAT	1385
Db	1734	TGATGCTGACCGCAGTGGCAGCTGACCTTCGGCCGACCTCAGCCATTTGTCACAGCT	1793
Qy	1386	---GCAGTACTGCACCAAGCTGTGTGTCACCGGGAAGCCCAAGGACAGATGGTGTGCCA	1442
Db	1794	GCGCCGCCCTGTGTGCTTGTGCTCTGTGCTGGCCACTCAATGGCCATGCCATGTGCCA	1853
Qy	1443	GACCCGCACTTCCCTGGGCGGATGGACACAGCTGTGGCAGGGGCAAGCTTCGCTCAA	1502
Db	1854	GACCAAACACTCGCCCTTGGGCGATGGACACCTTCGGGGGCCACAGGCCCTGCATGGG	1913
Qy	1503	AGGGGCTCGCTGGAGAGACAACTCTCAACAGCACAGGT-----GGATGTTCTCTG	1556
Db	1914	TGCTCGCTGCCTCCACATGGACAGCTCCAGACTTCAATATCCACAGCTGTGGCTG	1973
Qy	1557	GGCCAATGGGATCCCTATGGCCCTTGTGGGCACATGTGGTGGGGCGCTGCAGCTGGC	1616
Db	1974	GGTCTTGGGACCAATGGGCTGACTGCTCGGACCTGTGGGGGTGTGTCCAGTTCTC	2033
Qy	1617	CAGGAGGCAGTGCACCAACCCACCCCTGCCAACGGGGGCAAGTACTTCCAGGAGTGTG	1676
Db	2034	CTCCGAGACTGCAGAGCCCTGTCCCCGGNATGGTGGCAAGTACTGTGAGGGCCGCG	2093
Qy	1677	GGTGAATACCGATTCGTGAACCTGGAGCCCTGCCCGAGCTCAGCCTCCGGGAAGAGCTT	1736
Db	2094	TAGCCGCTTCCGCTCTCGAACACTGAGGACTGGCCAA---CTGGCTCAGCCCTGACCT	2150
Qy	1737	CCGGGAGGACGAGTGTGAGGCTTTCAACGGCTACAACACAGCACCAACCGGCTCACTCT	1796
Db	2151	CCGCGAGGAGAGTGTGTGCTGTACAAACCGCACCGACCTCTTCAAGAGCTTCCGAGG	2210
Qy	1797	CGCGTGGCATGGGTGCCCAAGTACTCCGGCTGTCTCCCGGGACAAAGTGCAGCTCAT	1856

Db	2211	GCCATGGACTGGGTTCCTCGCTACACAGGGCTGCCCCCCCAGGACCGTCAAACTCAC	227
Qy	1857	CTGCCAGGCAATGGCACTGGCTACTTCTATGTCTGCACCCCAAGGTGGTGACGGCAC	1916
Db	2271	CTGCCAGGCCCGGGCACTGGGCTACTACTATGTCTGGAGCCACGGGTGGTAGATGGAC	2330
Qy	1917	GCTGTGCTTCTCTGACTCAACCTCTCGTCTGATCCAAAGCAAGTGCATCAAGGCTGGCTG	1976
Db	2331	CCCTGTTCCTCCGGACAGCTCTCGGTCTGTCTCAGGGCCGATGCATCCATGTGGCTG	2390
Qy	1977	TGATGGGAACCTGGGCTCCAAGAAGAGATTGCACAAGTGTGGGTGTGTGGGGGAGACAA	2036
Db	2391	TGATCGCATATTGGCTCCAAGAAGAGTTTGACAAGTGCATGTGTCCGGAGGGGACGG	2450
Qy	2037	TAAGAGCTGCAAGAAGGTGACTTGGACTCTTCAACCAAGCCCATGCATGGCTACAAATTCGT	2096
Db	2451	TTCTGTGTTSCAGCAAGCAGCTCAGGCTCCTTCAGGAAATTCAGGTACGGATCAACAATGT	2510
Qy	2097	GGTGGCATTCCTCCGAGGGCCCTCAAGCATCAGCATCTCCGCAGCGCGGTTACAAGGGCT	2156
Db	2511	GGTCATATCTCCCGGGGGCCACCCACATTCTTCTCCGGCAGCAGGGAACCCCTGGCCA	2570
Qy	2157	GATCGGGGATGACAACCTACTGGCTCTGAAGAAACAGCCAAGCAAGTACCTGCTCAACGG	2216
Db	2571	CCGGAG-----CATCTACTTGGCCCTGAAGCTGCCAGATGGCTCCTATGCCTCAATGG	2624
Qy	2217	GCATTTGTTGGTGTCTGGCGGTGGAGGGGACCTGTGTGTGTT--AAGGGCAGTCTCC--TGCG	2273
Db	2625	TGAATACACGCTGATGCCCTCCCCACACAGATGTGGTACTGTCTGGGGCAGTCAGCTTGGC	2684
Qy	2274	GTACAGGGCACGGGCACACAGGGTGGAGGCTTCGAGGCTTCCCGGCCCATCTCTTGAGCC	2333
Db	2685	CTACAGCGGGGCCACTGCAGCCCTCAGAGACATGTTCAGGCCATGGGGCCCATCTGGCCAGCC	2744
Qy	2334	GCTGACCGTGGAGGTCCCTCTCCGTGGGGAAGATGACACCGGCCCGGGTCCCGCTACTCCTT	2393
Db	2745	TTTGACACTGCAGTCTCTAGTGGCTGGCAACCCCGAGGACACACGSCCTCCGATACAGCTT	2804
Qy	2394	CTATCTGCC	2403
Db	2805	CTTCGTGCC	2814

```

RESULT 7
US-09-122-126B-14
US-09-122-126B-14, Application US/09122126B
, Patent NO. 6451575
, GENERAL INFORMATION:
, APPLICANT: Bristol-Myers Squibb Company
, TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES
, FILE REFERENCE: DM6909
, CURRENT APPLICATION NUMBER: US/09/122,126B
, CURRENT FILING DATE: 1998-07-24
, NUMBER OF SEQ ID NOS: 21
, SOFTWARE: PatentIn version 3.0
, SEQ ID NO 14
, LENGTH: 3250
, TYPE: DNA
, ORGANISM: Homo sapiens
, FEATURE:
, NAME/KEY: CDS
, LOCATION: (121)..(2910)
US-09-122-126B-14

```

Query Match	19.2%;	Score 547.4;	DB 4;	Length 3250;
Best Local Similarity	57.7%;	Pred. No. 3.8e-110;		
Matches 1043; Conservative	0;	Mismatches 746;	Indels	18;

[illegible]

Db	944	CTGACGGCTCATGGCGCGGTTTGTATGGCGGGGGCTCGACATTACCTGCTGACCTGG	1003
QY	737	TGGCAACGGGGCGGCGGACTCTACCGGCATCCAGACATCTCAACCCCATCAACATCGTTG	796
Db	1004	CCTCCATCGCCAATAGGCTGTACAGCCATGCTTAGCATCGAGNACCACATCGCTGCGCG	1063
QY	797	TGGTCAAGGTGCTGCTTCTTAGAGATCGTGACTCCGGGCCCAAGGTACACGGCAATCGG	856
Db	1064	TGGTGAAGGTGGTGGTGTAGGCGCAAGACGAGAGCTGGAAGTGAGCAAGAAGCTG	1123
QY	857	CCCTGACGCTGCGCAACTTCTGTGCTGCGCAAGAAAGCTGNACAAAGTGAAGTCAAGC	916
Db	1124	CCACCACACTAAGAACTTTTGAAGTGGCAGCACAACAACACCGCTGGAGATGACC	1183
QY	917	ACCCGAGTACTGGGACACTGCCATCCTCTTCCACGAGCAGGACCTGTGTGGAGCCACCA	976
Db	1184	ATGAGGAGCACTACGATGCAGCTACCTCTTTACTCGGAGGATTTATGTGGGCATCAT	1243
QY	977	CCTGTGACACCTGGGATGGGTGATGTGGGTACCATGTGTGACCCCAAGAAAGCTGCT	1036
Db	1244	CATGTGACACCTTGGGAATGGCAGAGCTTGGACCATATGTTCTCCAGAGCGCAGCTGTG	1303
QY	1037	CTGTCAATTGAGGACGATGGCTTCCATCAGCCTTCCACACTGCCACGAGAGCTGGGCCAG	1096
Db	1304	CTGTGATTGAAGACGATGGCTTCCAGCAGCCTTCCACTGTGGCTCAGCAAAATCGGACAT	1363
QY	1097	TGTTCAACATGCCCATGACAATGTGAAGTCTGTGAGGAGGTGTTGGGAAGCTCGAG	1156
Db	1364	TACTTGGCCTCTCCCATGAGGATCCAAATCTGTGAAGACACTTTGGTTCACAGAAG	1423
QY	1157	CCAACACATGATGTCCGACCCCTATCCAGATCGACCGGTGCGCAACCCCTGTGACGCT	1216
Db	1424	ATAAGCGCTTAATGTCTTCATCCTTACAGCATTTGATGCATCTAAGCCCTGTGCTG	1483
QY	1217	GCAGTGTGTCATCATCACGACTTCTTGGACAGGGCGAGGTGACTGCTCTCTGGACC	1276
Db	1484	GCACCTCAGCCACCATCAGAAATCTCTGGATGATGGCCATGTTGTTCTGTGGACC	1543
QY	1277	AACCAGCAAGCCCATCTCCCTCGCCGAGGATCTGCCGGGCGGCAGCTACACCTCGAGCC	1336
Db	1544	TACCACGAAGACAGATCTCTGGGCCCGAAGAACTCCACGAGACACCTACGATGCCACCC	1603
QY	1337	AGCAGTCGAGCTGGCTTTTGGGTGGGTCCAGGCCCTGTCCTTACATGCATGACTGCA	1396
Db	1604	AGCATGCAACCTTGACATCTCGGGCCCTGAGTACTCGTGTGTCCGGCATGGATGTCTGTG	1663
QY	1397	CCAAGCTGTGGTGCACCGGAAGGCCAAGGACAGAGTGTGTGCCACCGGCCACTTCC	1456
Db	1664	CTGCGCTGTGGTGTGTGTGGTACCCAGGGCCAGATGGTGTGTGTGACCAAGAAGCTGC	1723
QY	1457	CCTGGGCCGATGSCACCACTCTGTGGCGAGGCAAGCTCTGCTCAAAAGGGCCTGCGTGG	1516
Db	1724	CTGCGGTGGAAGGAGCGCCTTGTGGAAGGGGAGAACTGCTGCTCAGGGCAAAATGTGG	1783
QY	1517	AGAGACACAACCTCAACAAGCACAGGGTG-----GATGGTTCCTGGGCCAAATGGGATC	1570
Db	1784	ACAAAACCAAGAAAAAATATTATCAAGCTCAAGCCATGGCAACTGGGGATCTTGGGAT	1843
QY	1571	CCTATGGCCCCCTGCTGCGCCACATGTGTGGGGCGTGTGACCTGGCCAGGAGCACTGCA	1630
Db	1844	CCTGGGGCCANGTTCCTCGCTCATGTGAGGAGGAGTGCASGTTTGGCCATTCGTCAC	1903
QY	1631	CCAACCCCACCCCTGCAACGGGGCAAGTACTGCGAGGAGTGAAGGTTGAAATCCGAT	1690
Db	1904	ATAACCTTCTCCAGAAACACGACCGCTACTGCACAGGAAGAGGGCCATCTACCGCT	1963
QY	1691	CCTGCAACCTGGAGCCCTGCCCCAGCTCAGCCTCCGGAAGAGCTTCCGGGAGGAGCACT	1750
Db	1964	CCTGCACTCTCATGCCCTGCC-----CACCCAATGTTAAATCATTTGTCATGAACAGT	2017
QY	1751	GTGAGCCTTTCACGGCTACACACACAGCACCCCGCTCTACTCTGCGCGTGGCAAGGG	1810

RESULT 8
US-09-369-364A-1
; Sequence 1, Application US/09369364A

```

: Patent No. 6391610
:
: GENERAL INFORMATION:
:
: APPLICANT: Apte, Suneel
:
: APPLICANT: Hurskainen, Tiina L.
:
: APPLICANT: Hirohata, Satoshi
:
: TITLE OF INVENTION: Nucleic Acids
:
: FILE REFERENCE: 26473/4007/10-30-00
:
: CURRENT APPLICATION NUMBER: US/09/3
:
: CURRENT FILING DATE: 1999-08-06
:
: NUMBER OF SEQ ID NOS: 31
:
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ ID NO 1
:
: LENGTH: 3002
:
: TYPE: DNA
:
: ORGANISM: mus musculus ADAMTS-5
:
: FEATURE:
:
: NAME/KEY: CDS
:
: LOCATION: (18)..(2810)
:
: US-09-369-364A-1

```

Query Match 19.1%; Score 545.8; DB 4; Length 3002;
Best Local Similarity 57.7%; Pred. No. 8.2e-110;
Matches 1042; Conservative 0; Mismatches 747; Index 18;

QY 617 GGTCTGGGCGCCAAAGCGTTTCTGCTCTATCCCGGGTACGTGGAGACGCTGGTGGTG 676
 DB 781 GGTGGAGGCGTAGGCGCGTTTCTATCTCCAGGCGCCGCCAGGTGGAGCTCTCTCTGGTGG 840
 QY 677 CGGACGAGTCAATGTGCAAGTTCCAGGGGCGGACCTGGAAATTTATCTGCTGAGCGTGC 736
 DB 841 CTGACTCGTCCATGGCCAGGATGATGGGGGCGCTCGAGCATTTACCTGCTGACCATGG 900
 QY 737 TGGCAAGGGGCGGCGACTCTACCGGCATCCCAAGATCTCAACCCCATCAACATCGTTG 796
 DB 901 CCTCCATCGCCAAAGCGTGTACAGTCAATGCAAGCATTTGAGAACCCACATCCCGCTGGCG 960
 QY 797 TGGTCAAGTGTCTCTTTAGAGATCTGACTCCGGGCGGACGCTACCGCAATGGCG 856
 DB 961 TGGTGAAGTGGTGGTCTGACGCGCAAGACACGAGTCTGGAGGTGACAAGATGGCG 1020
 QY 857 CCCTGAGCTGCGCAACTTCTGTGCTGCGAGGAAGCTGAACAAAGTGAAGTGAACG 916
 DB 1021 CCAGCAGCTCAAGAACTTTTGCAAATGGCAGCACCAACATAACCAAGCTAGGGGATGATC 1080
 QY 917 ACCCGAGTACTGGGACACTGCGCATCTCTTACAGGCGGAGGACCTGTGGAGCAACA 976
 DB 1081 ACAGAGAGCTACGATGCGACCATCTGTTCACCGGAGGAGTATTTATGTGGGCACTATT 1140
 QY 977 CCTGTGACACCTGGGCATGGCTGATGTGGTACCATGTGTGACCCCAAGAGAAGCTGCT 1036
 DB 1141 CATGTGACACCTGGGAATGGCAGAGCTTGGGACCATATGTCTCCGGAGCGAGCTGTG 1200
 QY 1037 CTGTGATGAGGAGGAGTGGGCTTCCATGACGCTTACACCTGCGCCAGAGCTGGGCGCAG 1096
 DB 1201 CAGTGATTGAAGATGATGGCTCCATGCGACCTTCACTGTGGCTCATGAATTTGGGCATC 1260
 QY 1097 TGTTCACATGCCCCATGACAAATGTGAAGCTGTGAGGAGGTGTTTGGGAAGCTCCGAG 1156
 DB 1261 TACTTGGCTTCTCATGAGCATTCANATCTGTGAAGAGAACTTCGGTACTACAGAAG 1320
 QY 1157 CCAACACATGATGTCCCGGACCTCATCCAGATGACCGGTGCCAACCCCTGGTACGCT 1216
 DB 1321 ACAAGCGTTAATGTCTTCAATCTTACCAGCATGATGATCCAAAGCCCTGGTCCAAAT 1380
 QY 1217 GCAGTGTGCAATCATCAGCATCTCTGACAGCGGCGAGTACTGCTCTCTGAGCC 1276
 DB 1381 GCAGTGTGAGCATCATCAGAAATCTGTGATGATGTGTAATTTGTTCTAGACC 1440
 QY 1277 AACCCAGAACCCCATCTCCCTGCGGAGGATCTGCGGCGGCGAGTACACACCTGAGCC 1336
 DB 1441 TACCACGGAGCAGATTTTGGTCCCAGGAACTCCAGGACAGCTAGCATGCCACCC 1500
 QY 1337 AGCAGTGTGAGTGGCTTTTGGGCTGGGCTCCAGCCCTGTCTTACATGCACTACTGCA 1396
 DB 1501 AGCAGTGTGAGTGGGCTTGGGCTGAGTACTCGGTGCTGCGCTGGCATGGATGCTGTG 1560
 QY 1397 CCAAGCTGTGGTGCACCGGAGGCAAGGACAGATGGTGTGCCAGACCGCCACTTCC 1456
 DB 1561 CGCGCTGTGGTGTGCTGTGGTGGCCCAAGGCAAAATGGTGTGCTGTACCAAGAACTGC 1620
 QY 1457 CCTGGGCGGATGGCACCAGCTGTGGGCGGCAAGCTCTGCTCAAGGGGCGCTGGGTGG 1516
 DB 1621 CGGCTGTGGAGGCACTCCCTGTGGGAGGGAAGAGTCTGCCITCAAGGCAAAATGTGG 1680
 QY 1517 AGAGACACAACTCAACAGCAGAGGTG-----GATGGTTCCTGGGCGAAATGGGATC 1570
 DB 1681 ACAAACTAAGAAAAAATTTACTCGACATCAAGCCATGGAATTTGGGGTCTCTGGGGCC 1740
 QY 1571 CCTATGGCCCTGTCCGCGACATGTGTGGGCGGTGAGTGGCGGAGGAGGAGTGA 1630
 DB 1741 CTTGGGTTCAGTGTCTCGCTCATGCGGGGAGGAGTGTGCTACCGCCATGTA 1800
 QY 1631 CCAACCCACCCCTGCCAAAGGGGCAAGTACTGCGAGGAGTGAAGGTGAATACCGAT 1690
 DB 1801 ATACCCCTGCACCTCGAAACAGTGGCCGCTACTGACAGGGAAGGGCCATATACCGTT 1860

QY 1691 CTTGCAACCTGGAGCCCTGCCCCAGCTAGCCCTCCGGAAAGAGCTTCCGGGAGAGCAGT 1750
 DB 1861 CTTGCAAGTGTACACCCCTGCC-----CACCAATGTAAATCTTTTCGGCATGAGCAGT 1914
 QY 1751 GTGAGGCTTTCAAGGGCTAGAACACACAGCACCACCGGCTCACTCTCGCCGTGCGCATGGG 1810
 DB 1915 GTGAAGCCCAAAATGGCTATCAGTCTGATGCANAAAGGAGTCAAAACATTTTATGAATGGG 1974
 QY 1811 TGCCCAAGTACTCTCGGCGGTCTCCCGGGACAAAGTGCAGCTCATCTGCGGAGCCCAATG 1870
 DB 1975 TTCCCAATATGACAGGTCTCTCGCGGACAGTGTGCAAGCTTACCTGCGAGAGCTAAG 2034
 QY 1871 GCATGAGCTACTTCTATGTGCTGGCACCCAAAGTGTGGAGCGGACGCTGTGCTCTCTG 1930
 DB 2035 GCACAGGCTACTATGTGCTCTTTTCTCCAAAGTTAAGGATGGGACTGAATGCGAGGCGT 2094
 QY 1931 ACTCCACCTCTGCTGTGCTCCAAAGCAAGTGCATCAAGGCTGCTGTGATGGGAACCTGG 1990
 DB 2095 ACAGCAACTGTGTGTCTCGGAGCGGTGTGTGAGAACTGGATGTGACGGCATTTATG 2154
 QY 1991 GCTCCAAAGAAAGAGATTTCGACAAAGTGTGGGTGTGTGGGGGAGACAATAAGAGCTGCAAGA 2050
 DB 2155 GCTCAAGCTCAATATGCAAGTGTGGAGTGTGGAGGGGATAAATCCAGTTGTACAA 2214
 QY 2051 AGGTGACTGGACTCTTCAACAGCCCATGCTAGCTACAAATTTGCTGTGCGCCATCCCG 2110
 DB 2215 AGATTATCGGAACCTTCAATAAAAAAGCAAGGTTTACTGACGTGTGTGAGGATCCCTG 2274
 QY 2111 CAGGCGCTCAAGCATCGACATCCGCCAGCGGTTTACAAAGGCGTATCGGGGATGACA 2170
 DB 2275 AAGGAGCAACCCACATAAAGTCCGACAGTTCAAGCCAAAGAGACAGACTAGATTCCCTG 2334
 QY 2171 ACTACTGGCTGTGAAGAACGCCAAGCAAGTACTCTCTCAAGGGGCAATTTCTGTGGTGT 2230
 DB 2335 CTTACTTAGCCCTGAAGAAGAACTGGCGAGTACTTTATCAATGGCAAGTACATGATTT 2394
 QY 2231 CGGCGTGTGGGCGGACCTGGTGTGAAGGCGAGTCTCTGGGTACAGCGGCGACGGCA 2290
 DB 2395 CCACCTTCAGAGACCATCATCGACATCAATGGTTCGCTCATGAATACAGTGGATGGAGCC 2454
 QY 2291 CAGCGTGTGAGAGCCCTGCA-----GGCTTCCCGGCCCATCTCTGAGCGGCTGACCCGTG 2344
 DB 2455 ACAGAGATGATTTTTTACATGGGATGGGCTATTACAGCCACAAAGAAATCTGTATCGTGC 2514
 QY 2345 AGGTCTCTCTCGTGGGAGAGTACACCGCCCGGCGCTGCTCTCTTCTATCTGCCCCA 2404
 DB 2515 AGATCTTGGCCACAGACCCCACTAAAGCGCTAGGCGCTCCGTTACAGCTTTTTTTGTTCCCA 2574
 QY 2405 AAGAGCC 2411
 DB 2575 AGAAGAC 2581

RESULT 9
 US-09-369-364A-14
 ; Sequence 14, Application US/09369364A
 ; Patent No. 6391610
 ; GENERAL INFORMATION:
 ; APPLICANT: Apte, Suneel
 ; APPLICANT: Hurskainen, Tiina L.
 ; APPLICANT: Hirohata, Satoshi
 ; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
 ; FILE REFERENCE: 26473/4007/10-30-00
 ; CURRENT APPLICATION NUMBER: US/09/369,364A
 ; CURRENT FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 31
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 14
 ; LENGTH: 2625
 ; TYPE: DNA
 ; ORGANISM: Mus musculus ADAMTS-9
 ; FEATURE:
 ; NAME/KEY: CDS

LOCATION: (2)...(2623)
US-09-369-364A-14

Query Match 16.5%; Score 471.4; DB 4; Length 2625;
Best Local Similarity 56.5%; Pred. No. 1.2e-93;
Matches 1009; Conservative 0; Mismatches 751; Indels 27; Gaps 6;

Qy 629 CCAAGCGTTTCGTCTATCCCGCGTACGTGGAGACGCTGGTGGCGGACAGTCAA 688
Db 372 CCAAGCGTTTCGTCTATCCCGCGTACGTGGAGACGCTGGTGGCGGACAGTCAA 431
Qy 689 TGGTCAAGTTCACGCGCGGACGCTGGAGACATATCTGCTGACGCTGCTGCAACGGCGG 748
Db 432 TGGTCAAGTTCACGCGCGGACGCTGGAGACATATCTGCTGACGCTGCTGCAACGGCGG 491
Qy 749 CGCGACTTCACGCGCGGACGCTGGAGACATATCTGCTGACGCTGCTGCAACGGCGG 808
Db 492 CTTCTATCTATAAGACTCAAGTATGGAAATTTAATTAATTTAATTTAATTTAATTTA 551
Qy 809 TGGTCTTAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 868
Db 552 TGGTCTTAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 611
Qy 869 GCAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 928
Db 612 AGAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 664
Qy 929 GGGACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 985
Db 665 --GACACAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
Qy 986 CCCTGGGATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1045
Db 723 CTTAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 782
Qy 1046 AGGACGATGGCTTCCATCAGCTTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1105
Db 783 AAGACGATGGCTTCCATCAGCTTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 842
Qy 1106 TGCCCATGACATGTGAAGTCTGTGAGGAGTGTGTTGGGAGCTTCCGAGCAACCA 1165
Db 843 TGCCCATGACATGTGAAGTCTGTGAGGAGTGTGTTGGGAGCTTCCGAGCAACCA 899
Qy 1166 TGATGCTCCGACCTCATCCAGATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1225
Db 900 TCATGGCACCAACACTGAATCTTACACCAACCTGCTGCTGCTGCTGCTGCTGCTGCTG 959
Qy 1226 CCATCATCAGGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1282
Db 960 AATCATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1019
Qy 1283 GCAAGCCATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1342
Db 1020 CCAGGACCTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1079
Qy 1343 GCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1402
Db 1080 GTGAAGTATGTTGGGCGAGCTCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1139
Qy 1403 TGTGCTGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1462
Db 1140 TGTGCTGCAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199
Qy 1463 CCGATGGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1522
Db 1200 CAGATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1259
Qy 1523 ACAACCTCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1582
Db 1260 AAA---TGGAGGCGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1316
Qy 1583 GCTCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1642

Db 1317 GCTCAAGAAAGCTGTGGAGGAGCATCAAAACAGCCATCAGAGAGTCCACAGACAGAGC 1376
Qy 1643 CTGCCAACGGGCAAGTACTCGAGGAGTGGAGGTGAAATACCATCTCTGCAACCTGG 1702
Db 1377 CAAAATGTTGGAGTACTGTGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1436
Qy 1703 AGCCCTGCCCGACCTCAGCCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1762
Db 1437 AGCCCTGCATGAAGCAAGAGCGAG-----ACTTCGAGAGGAGGAGGAGGAGGAGGAG 1490
Qy 1763 AGGCTACACACACAGCACCAGCTCAGCTCGCCGCTGATGCTGCTGCTGCTGCTGCTGCTG 1822
Db 1491 ATGCAAAACATCTCAACATCAATGGTGTGCTGCCAGCTGACGCTGCTGCTGCTGCTGCTG 1550
Qy 1823 CCGCGTGTCTCCCGGCAAGTCAAGCTCATCTGCCAGGCAATATGCACTGCTGCTGCTGCTG 1882
Db 1551 GCGGAATTTGATGAAGGAGCGGTGCAAGTGTCTGAGAGTGGCAGGAGGAGGAGGAGGAG 1610
Qy 1883 TCTATGCTGGACACCAAGGCTGGAGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1942
Db 1611 ACTACCAAGCTCCGAGACAGAGTATTGACGGAACCCCTTGTGGCCAGGACACAAATGACA 1670
Qy 1943 TCTGTGTCGAAGCAAGTGCATCAAGCTGGCTGTGATGGGAAACCTGGGCTGCCAAGA 2002
Db 1671 TCTGTGTCGAAGCTTTCGCGCAAGCTGGATGATCATATTTAAACTCAAGGTC 1730
Qy 2003 GATTGCAAGTGTGGGCTGTGGGAGGAGCAATTAAGAGCTGCAAGAGGAGTGAAGTGA 2062
Db 1731 GGAAGATAAATGTGGATTTGTGTGGAGTAACTTCTTCATGCAAAACAGTGGCAGGAA 1790
Qy 2063 TCTTACCAAGCCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2122
Db 1791 CATTTACACTGCTCATATGTTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1850
Qy 2123 GCATGACATCCCGCAGCGCTTACAAAGGCTGATCGGGGATGATCACTACCTGGCTC 2182
Db 1851 GCATTGAGTGGTGCAGCAGCTTCTCAGGGAAGTCTGAGGATGATCACTACCTAGCTT 1910
Qy 2183 TGAAGACACCAAGCAAGTACTGCTCAACGGGATTCGCTGCTGCTGCTGCTGCTGCTGCTG 2242
Db 1911 TATCAACAGTAAATGTGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1970
Qy 2243 GGGACCTGGTGTGAAGGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2302
Db 1971 GGGAGTCCCGCTGGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2030
Qy 2303 GCCTGAGGCTTCCCGGCGCATCTGAGCGGCTGAGCGGCTGAGCGGAGGAGTCTGCTGCTG 2362
Db 2031 GACTGAAGTGTACGAGCGCTATCGAGGAGGAGTCTTCTCTTCTGCTGCTGCTGCTGCTG 2090
Qy 2363 AGATGACACCGCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2409
Db 2091 AGCTGATACCCAGATGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2177

RESULT 10
US-09-369-364A-12
; Sequence 12 Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369, 364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 5804
; TYPE: DNA
; ORGANISM: Homo sapiens ADAMTS-9

1572 GGCAGTGTGAGGACCAAGTGTGGGGTGTGTGGGGGA 2031

Db 235 TGATCGTAGAAGATGAAATATGGGGCCAGAGGTGTCGACATGGGGGCTTACACTGC 294
 Qy 869 GCACTTCTGTGCTGGCAGAGAGCTGAACAAGTGAAGTGAACAAGCAGCCCGAGTACT 928
 Db 295 GTAACTTCTGCACTGGCAGGGGGTTCACACAGCCCGACCGCCAGACACT 354
 Qy 929 GGGACACTGCATCTCTTCCACGAGGAGGAGCTGTGTGG- --AGCCACACCTGTGACA 985
 Db 355 ACGACAGGCACTCTCTCTCACCAGACAGAACTTCTGTGGGCGAGGAGGCTGTGTGACA 414
 Qy 986 CCCTGGGCACTGGTGTGATGTGGGTACCATGTGTACCCCAAGAGAGCTGTGTCTGTG 1045
 Db 415 CCCTGGGCTGTGGCAGACATCGGACCACTTGTGACCCCAAGAGCTGTGTGTGATCG 474
 Qy 1046 AGGACGATGGCTTCCATCAGCTTCCACACTGCCACAGAGCTGGGCGACGCTTCAACA 1105
 Db 475 AGGATAGGGGCTCCAGGCGCCACACCTTGGCCCATGAACCTAGGCGAGCTCTCAGCA 534
 Qy 1106 TGCCCCATGACAATGTGAAAGTCTGTGAGGAGGTGTTTGGGAGAGCTCCGAGCCAAACACA 1165
 Db 535 TGCCCCAGAGACTTCCAAAGCCCTGACACAGGCTCTTCGGGCGCATGGGCAAGCAGCAG 594
 Qy 1166 TGATGTCCCGACCTCATCCAGATCAGCTGCCAACCCCTGTGTCAGCTGCAGTGTGCTG 1225
 Db 595 TGATGGCAGCTGTGTCACCTGAACAGAGAGCTGCCCTGTGTCCTCCCTGAGCGCCA 654
 Qy 1226 CCATCATCAGCAGCTTCTTGGACAGCGGG 1254
 Db 655 TGTTCAGGGCTGCACCTGCAGGGTGG 683

RESULT 13

US-09-392-184-5/c

; Sequence 5, Application US/09392184

; Patent No. 6395889

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN

; TITLE OF INVENTION: PROTEASE HOMOLOGS

; FILE REFERENCE: 5800-55

; CURRENT APPLICATION NUMBER: US/09/392,184

; CURRENT FILING DATE: 1999-09-09

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 5

; LENGTH: 5357

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(5357)

; OTHER INFORMATION: reprotolysin (ADAM family of metalloprotease)

; NAME/KEY: misc_feature

; LOCATION: (1)...(5357)

; OTHER INFORMATION: n = A,T,C or G

US-09-392-184-5

Query Match 8.7%; Score 247.2; DB 4; Length 5357;
 Best Local Similarity 51.8%; Pred. No. 7.5e-45;
 Matches 927; Conservative 0; Mismatches 778; Indels 83; Gaps 13;

Qy 630 CAAGGTTTCTGTCTATCCCGGGTACGTGGAGACGCTGTGTGTCGGGAGAGTCAAT 689
 Db 4617 CCAGCGGTGCTGACGAAGAAGAGTGGGTGGAGACCTTGGTAGTGTATGCCAAAT 4558
 Qy 690 GGTCAAGTTTCCAGG- --CGCGACCTTGAACATTATCTGTGACGCTGTGTGGAACGGC 746
 Db 4557 GGTGAGTATACCGAGACGCGCAGGTGTGAGAGCTATGTGTGACCATCATGAACATGGT 4498
 Qy 747 GCGCGACTCTACCGCCATCCAGCATCTCAACCCCATCAACATCTGTGTGTCAGAGT 806
 Db 4497 GGTGCTGCTTTCATGACCCAGCATTTGGGAACCCCATCCACATCACCATTGTGCGCT 4438

Qy 807 GCTGTTCTTAGAGATCGTGTACTCGGGGCCCAAGGTCAACGGCAATATGGGGCCCTGACGCT 866
 Db 4437 GGTCTGCTGGAAATGAGGAGGAGGAGCTAAAGATCAACGACCATGAGAGCAACACCT 4378
 Qy 867 GCGCAACTTCTGTGCTGGCAGAGAAAGCTGAACAAGTGAAGTGAACAAGCAGCCGAGTA 926
 Db 4377 GAAGAGCTTCTCAAGTGGCAGAAAGCATCAACATGAAGGGGATGCCATFCCCTGCA 4318
 Qy 927 CTGGGACACTGTCATCTCTTCCACAGGACGAGACCTGTGTGGAGCCACCA- - - - -CCTG 980
 Db 4317 CCATGACACTGTCATCTCTGCTCACCAGAAAGGACCTGTGTGACGCAATGAACCGGCCCTG 4258
 Qy 981 TGACACCTGGGCGATGTGTGTGGGTACCATGTGTGACCCCAAGAGAAAGCTGCTCTGT 1040
 Db 4257 TGAGACCTTGGGACTGTCCCATGTGGCGGATGTGCCAGCCGACCCAGCATGACGAT 4198
 Qy 1041 CATTGAGGACGATGGCTTCCATCAGCTTCCACACTGCCACAGCTGGGCGACGCTGT 1100
 Db 4197 CAACGAGNACAGGGCTTGGCTTCTACTGTAGCCCAAGAGCTCGGCGACAGTTT 4138
 Qy 1101 CAACATGCCCCATGACAATGTGAAAGTCTGTGAGGAGGTGTTTGGGAAAGCTCCCGAGCAA 1160
 Db 4137 TGGCATTCAGCATGACGGAAGGCGCAATGACTGTGAGGCCGTTGGGAAACGAC- - - - -C 4084
 Qy 1161 CCACATGATGTCCCGACCTCATCCAGATCGACCTGCAACCCCTGTGTAGCCTGAG 1220
 Db 4083 TTTTCATGATGTCCACAGCTCTGTAGACCCGCTCCCTCAGCTGTGTGCTGAG 4024
 Qy 1221 TGCTGCATCATCAGCGACTTCTTGGACAGCGGCGAGGTGACTGCTCTCTGGACCAACC 1280
 Db 4023 CGGCCAGTATATCACCAGGTTCTTGGCCGTGGGTGGGCGCTGTGCTTGGAGACCTCC 3964
 Qy 1281 CAGCAAGCCC- - - - -ATCTCCTGCCGAGGATCTGCGGGCGCCAGCTACACCTGAGCA 1337
 Db 3963 TGCCAAAGGACATATCGACTTCCCTCGGCTGCACTGCGGCTCTCTATGATGAAGCA 3904
 Qy 1338 CGAGTGCAGAGCTGGCTTTTGGGTGGGCTCCCAAGCCCTGTCTCTTACATGACGTA- - - - -CTG 1394
 Db 3903 CGAGTGCCTGCTCCAGTACGGGCTTACTTCTGCGAGGACATGGATAATGTCTG 3844
 Qy 1395 CACCAAGCTGTGTGACCGGGAAGGCCAAGGAGACAGATGTGTGCCAGACCCGCACTT 1454
 Db 3843 CCACACACTCTGTGTCTGTGGGACCACTGTCTACTCCAAAGCTGGATGCAAGCC- - - - - 3788
 Qy 1455 CCCTGGGCGGATGGCAGCAGCTGTGGGAGGCGAGCTGCTCCCTCAAGGGGCTGGT 1514
 Db 3787- - - - -GTGGAGCGCACCGGTGTGGGAGAAATAGTGGTGTCTCAGTGGGAGTGGT 3735
 Qy 1515 GGAGAGACACAACTCAACAAGCACAGGGTGGATGGTTCCTGGGCAATAGSGATCCCTA 1574
 Db 3734 ACCGTGGGCTTCCGGCCCGAGGCC- - - - -GTGGATGGTGGCTGTGGCTGGAGCGCTG 3678
 Qy 1575 TGCCCTCTGCTGCGCAGATGTGGTGGGCGGTGACGTGGCGCAGGAGGACGTGGACCA 1634
 Db 3677 GTCCATCTGCTACGGAGCTGTGGCATGGGCTGTACAGACGCGCGAGCGACGTACGCA 3618
 Qy 1635 CCCCACCTTGCACACGGGGCAAGTACTGCGAGGAGGTGAGGTTGAAATACCGTCTG 1694
 Db 3617 GCCTACGCCCAATACAAAGGAGATACTGTGTGGGTGAGCGCAAGCGCTTCGCGCTCTG 3558
 Qy 1695 CAACCTGGAGCCCTGCCCGCAGCTACGCTCCGGAAGAGCTTCCGGGAGGAGCATGTGA 1754
 Db 3557 CAACCTGCAGGCTGCC- - - - -CTGCTGCCGCCCTCTCTCCGCGACGTCCAGTGCAG 3504
 Qy 1755 GGTCTTCAACGGGTACACACAGCAGCAACCGGCTACTCTGCGCGTGGCATGGGTGCC 1814
 Db 3503 CCACCTTTCAGCTATGCTCTACAAAGGCGCAGCTGCA- - - - -CATGGTGGC 3456
 Qy 1815 CAAGTACTCGGCGTCTCTCCCGGAGCAAGTGCAGCTCATCTGCGGAGCAATGGCAC 1874
 Db 3455 CGTGGTCAATGAGCTGAACCCC- - - - -TCGAGCTGCACTGCCGCGCGCGNATGA 3405
 Qy 1875 TGGCTACTTCTATGTGTGCGCACCCCAAGGTGTGTGGACGCGCATGTGTCTCTCTGACTC 1934

Db 3404 GTACTTTGCCGAGAAGCTCGGGACGCCGTGGTGCATGGGCACCCCTGCTACCAAGTCCG 3345
 QY 1935 CACCTCCG-----TCGTGTCCAGCAAGTGCATCAAGCGTGGCTGTGATGGGAACCT 1988
 Db 3344 AGCCAGCGGGACCTTTCATCAACGGCATCTGTAGAACGTGGGTGTGACTTCGAGAT 3285
 QY 1989 GGGTCCCAAGAGAGATTCGACAAAGTGTGGGTGTGGGGGAGACAATAAGAGCTGCAC 2048
 Db 3284 TGACTCCGGTCTATGAGGACCGCTGTGGTGTGGCCAGGCAACGGTCCACCTGCACA 3225
 QY 2049 GAAGTGACTGGACTTCACCAAGCCCATGCATGCATCAACAATTCGTGGTGCCCATCCC 2108
 Db 3224 CACCGTGAGCGGACCTTCAGGAGGCGGAGGCGCTGGGTGTGATGTGGGTGGCTGAT 3165
 QY 2109 GCGAGGCGCTCAAGCATCAGCATCCGACGCGCGGTACAAAGGCTGATCGGGATGA 2168
 Db 3164 CCCAGCGGGCA-----CGGAGATCCCGATCCAAAGAGGTTGCCGAGGCTGC 3117
 QY 2169 CAACTACCTGGCTGTGAAGAACAGCCAAAGCAAGTACCTGCTCAACGGGCAATTCGTGGT 2228
 Db 3116 CAACTTCTTGSCATCGGGAGTGAGGACCCGAGAGTACTTCTCAATGGTGGCTGGAC 3057
 QY 2229 GTCGGCGGTGAGCGGACCTGTGTGTAAGGGAGTCTGCTCGGTGACAGCGGACGGG 2288
 Db 3056 CATCCAGTGAAGCGGACTTACAGGTGGCAGGGACACCTTCACATACGACGCGAGGG 2997
 QY 2289 CACAGCGGTGAGAGCTCGAGGCTTCCCGCGCCATCCTGGAGCGCTGACCTGGAGGT 2348
 Db 2996 CA---ACTGGGAGAACCTACGTCCTCCCGGTCCCAAGAGGCGCTGTGGATCCAGT 2940
 QY 2349 CCTCCGTGGGGAAGATACACCGCCCGGGTCCGCTACTCCTCTTA 2396
 Db 2939 GCTGTCCAGGAGAACCTCGGGGTGCATAGGATACACCATCA 2892

RESULT 14

US-09-369-364A-6
 ; Sequence 6, Application US/09369364A
 ; Patent No. 6391610
 ; GENERAL INFORMATION:
 ; APPLICANT: Apte, Suneel
 ; APPLICANT: Hurskainen, Tiina L.
 ; APPLICANT: Hirohata, Satoshi
 ; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
 ; FILE REFERENCE: 26473/4007/10-30-00
 ; CURRENT APPLICATION NUMBER: US/09/369,364A
 ; CURRENT FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 31
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 3218
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens ADAMTS-7
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (13)..(3003)
 us-09-369-364A-6

Query Match 8.4%; Score 240.2; DB 4; Length 3218;
 Best Local Similarity 51.4%; Pred. No. 2.2e-43;
 Matches 933; Conservative 0; Mismatches 798; Indels 84; Gaps 13;
 QY 562 CTGGACCTTTACAGCGCGCGGGGCTTCGGGAGAGTTCGTAGCGGGCAGGTCT 621
 Db 646 CTGGAGTCTCGAGCGGCGCTTGGGAGCAGCGGACGAGTGGCGGCGCACCGCTGAGG 705
 QY 622 GGGCGGCCAAGCGTGTGCTATCCCGCGGTACGTGGAGACGCTGTGTGCGGGAC 681
 Db 706 CGTCTACACCGCGGTGGTFCAGCAAGAGAGTGTGTGAGACCTGTGTAGTAGTAT 765
 QY 682 GAGTCAATGCTCAAGTTCACGG---CGCGGACCTGGACATATATCTGTGAGCTGCTG 738

Db 766 GCCAAAATGGTGGAGTACCACGACGCGCAGGTTGAGAGCTATGTGCTGACATCATG 825
 QY 739 GCAAGCGCGCGGAGTCTTACCGCCATCCAGCATCTCCAAACCCCATCAACATCTGTGTG 798
 Db 826 AACATGGTGGCTGGCGCTGTTCATGACCCAGCATTTGGAGCCCATCCATCAACATT 885
 QY 799 GTCAAGGTGCTCTTCTTAGAGATCGTACCTCCGGGCCAAGGTCAACGGCAATCGCGC 858
 Db 886 GTGGCGCTGGTCTGTGGAGATGAGGAGGAGACCTAAAGATCAGGACCATGACAG 945
 QY 859 CTGAGCTGCGCAACTTCTGTGCTGGCAGAGAAAGCTGAACAAAGTGAAGTACAAAGC 918
 Db 946 AACACCTCAAGAGCTTCTGCAAGTGGCAGAAACATCAACATGAAGGGATGCCCAT 1005
 QY 919 CCGAGTACTGGGACACATGCTCTCTTACACGAGGAGACCTGTGTGGAGCCACCA-- 976
 Db 1006 CCGCTGCACCATGACATGCCATCTCTGCTACACGAAAGGACCTGTGTGAGCCATGAC 1065
 QY 977 ---CCTGTGACACCTTGGGCTGCTGTGTGGGTACCATGTGTGACCCCAAGAGAAGC 1032
 Db 1066 CCGCCCTGTGAGACCTTGGGACTGTCCCATGTGGCGGGCATGTGCCAGCCGACCGCAGC 1125
 QY 1033 TGCTCTGTCAATTGAGGACGATGGGCTTCCATCAGCCTTCAACACTGCCACGAGTGGGC 1092
 Db 1126 TGACGATCAACGAGGACACGCGCTGCGCTGGCTTCACTGTAGCCACGAGCTCGG 1185
 QY 1093 CACGTGTTCAACATGCCCATGACAAATGTGAAGTCTGTGAGGAGGTGTTTGGGAGCTC 1152
 Db 1186 CACAGTTTGGCATTCAGCATGACGGAAGCGCAATGACTGTGAGCCCGCTTGGGAACGA 1245
 QY 1153 CGAGCAACACATGATGTCCCGACCTTCCATCCAGATCGACCGTGCCAAACCCCTGGTCA 1212
 Db 1246 CTTTTCATC-----ATGCTCCACAGCTCTGTAGAGCGGCTCCCTCACTCTGGTCC 1299
 QY 1213 GCTCGAGTGTGCCATCATCAGCTTCTTGGACAGCGGCGACGGTGAATGCTCTCTG 1272
 Db 1300 CGTGCAGCGCGCAGTATATACCAAGTTCTTGTACCGTGGGTGGGCTGTGCTGGAC 1359
 QY 1273 GACCAACACGACGAGCC---ATCTCCCTGCCGAGGATCTCCGGGCGCAGCTACACC 1329
 Db 1360 GACCTCTTGCACAGGACATTCAGCTTCCCTCGGTGGCCACCTGGGCTCTCTATGAT 1419
 QY 1330 CTGAGCAGAGTGGGAGTGGCTTTTGGCGTGGCTCCAAAGCCCTGCTTACATGAG 1389
 Db 1420 GTAAGCCACAGTGGCGCTCCAGTACGGGGCTACTCTGCTTCTGCGAGGACATGAT 1479
 QY 1390 TA---CTGACCAAGCTGTGTGACCGGGAAGGCAAGGACAGATGTTGTGCGAGACC 1446
 Db 1480 AATGTGCGCACACACTCTGTGTCTGTGGGGACCA-----CCTGTCACTGCC 1527
 QY 1447 CGCCACTTCCCTGGCGCATGCGCACAGCTGTGGCAGGCGCAAGCTCTGCTCAAGGG 1506
 Db 1528 AAGCTGGATGAGCTGTGGACCGGACCGGCTGTGGGAGAAATAAGTGTCTCTAGTGG 1587
 QY 1507 GCTCGCTGGAGAGACAACTTCAACAGCACAGGCTGGATGTTCTTGGGCCCAATGG 1566
 Db 1588 GAGTCGTACCGTGGGCTTCCGGCCGAGGCC---GTGGATGGTGGCTGTGCTGGCTGG 1644
 QY 1567 GATCCCTATGGCCCTGTCTGGGACATGTGGTGGGGCGGTGACGCTGCCAGAGGAGCAG 1626
 Db 1645 AGCGCTGTGCTCATGCTGCTACGGAGCTGTGGCATGGGCGTACAGAGCGCGGAGCGG 1704
 QY 1627 TGCACCAACCCCGCTGCCAAGCGGGGCAAGTACTCGAGGAGTCAAGGTGAGGTGAAATAC 1686
 Db 1705 TGCACGAGCTTACGCCCAATACAAAGGACATGACTGTGTGGTGGAGCGCAAGCGCTTC 1764
 QY 1687 CGATCTCTCAACTGGAGCCCTTGGCCAGCTCAGCTCCCGGAAAGAGCTTCCGGGAGGAG 1746
 Db 1765 CGCCTCTCAAGCTGCAGGCGCTGCC-----CTGCTGSCCGCCCTCTCTCCGCCAGCTC 1818
 QY 1747 CAGTGTGAGGCTTTCACAGGGCTACAAACACAGCAGCAGGCTCACTCTCGCGGTGGCA 1806
 Db 1819 CAGTGCAGCCACTTGTGAGCGCTATGCTCTACAAGGGCCAGCTGCACA-----CA 1866

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 19:41:41 ; Search time 339 Seconds
(without alignments)
9944.440 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgcttttgctgggcatact.....gcgtcctgagccgctgctga 2853

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues
Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2846.6	99.8	2853	10	US-09-965-631-3
2	2846.6	99.8	3446	10	US-09-965-631-7
3	2296.4	80.5	2469	9	US-10-163-316-3
4	2296.4	80.5	2940	9	US-10-163-316-1
5	1088	38.1	1104	10	US-09-965-631-5
6	955.8	33.5	966	10	US-09-965-631-1
7	636	22.3	2184	9	US-10-097-597-13
8	636	22.3	2184	9	US-10-097-580-13
9	636	22.3	2184	10	US-09-445-023A-13
10	628.8	22.0	2184	9	US-10-097-580-2
11	628.8	22.0	2184	10	US-09-445-023A-2
12	628.2	22.0	4676	12	US-10-105-929-1
13	628.2	22.0	3638	10	US-09-918-171A-8
14	592	20.8	4407	9	US-10-174-590-351
15	583.2	20.4	4407	9	US-10-176-758-351
16	583.2	20.4	4407	9	US-10-175-737-351
17	583.2	20.4	4407	9	US-10-173-706-351
18	583.2	20.4	4407	9	US-10-173-697-351
19	583.2	20.4	4407	9	US-10-173-697-351

20	583.2	20.4	4407	9	US-10-175-752-351	Sequence 351, App
21	583.2	20.4	4407	9	US-10-176-482-351	Sequence 351, App
22	583.2	20.4	4407	9	US-10-176-757-351	Sequence 351, App
23	583.2	20.4	4407	9	US-10-176-913-351	Sequence 351, App
24	583.2	20.4	4407	9	US-10-180-552-351	Sequence 351, App
25	583.2	20.4	4407	9	US-10-180-557-351	Sequence 351, App
26	583.2	20.4	4407	9	US-10-173-700-351	Sequence 351, App
27	583.2	20.4	4407	9	US-10-174-572-351	Sequence 351, App
28	583.2	20.4	4407	9	US-10-174-579-351	Sequence 351, App
29	583.2	20.4	4407	9	US-10-174-582-351	Sequence 351, App
30	583.2	20.4	4407	9	US-10-174-588-351	Sequence 351, App
31	583.2	20.4	4407	9	US-10-175-739-351	Sequence 351, App
32	583.2	20.4	4407	9	US-10-175-740-351	Sequence 351, App
33	583.2	20.4	4407	9	US-10-175-743-351	Sequence 351, App
34	583.2	20.4	4407	9	US-10-176-488-351	Sequence 351, App
35	583.2	20.4	4407	9	US-10-176-492-351	Sequence 351, App
36	583.2	20.4	4407	9	US-10-176-747-351	Sequence 351, App
37	583.2	20.4	4407	9	US-10-176-750-351	Sequence 351, App
38	583.2	20.4	4407	9	US-10-176-985-351	Sequence 351, App
39	583.2	20.4	4407	9	US-10-176-987-351	Sequence 351, App
40	583.2	20.4	4407	9	US-10-176-991-351	Sequence 351, App
41	583.2	20.4	4407	9	US-10-176-992-351	Sequence 351, App
42	583.2	20.4	4407	9	US-10-176-993-351	Sequence 351, App
43	583.2	20.4	4407	9	US-10-184-658-351	Sequence 351, App
44	583.2	20.4	4407	9	US-10-173-695-351	Sequence 351, App
45	583.2	20.4	4407	9	US-10-173-697-351	Sequence 351, App

ALIGNMENTS

RESULT 1
US-09-965-631-3
; Sequence 3, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddele, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; PRIOR FILING DATE: 2001-09-27
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2853
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-965-631-3

Query Match	99.8%	Score 2846.6;	DB 10;	Length 2853;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 2849;	Conservative	0;	Mismatches	4;
			Indels	0;
			Gaps	0;
Qy	1	ATGCTTTTCTGGGCATCCTAACCCCTTTCGCGGGGGAACCGCTGGAGGCTTTGAG	60	
Db	1	ATGCTTCTGCTGGGCATCCTAACCCCTTTCGCGGGGGAACCGCTGGAGGCTTTGAG	60	
Qy	61	CCAGAGCGGGAGGTAGTGGTTCCTCCGACTCGGAGCTCAGGACTCATTTTTCAGATCAGACATTT	120	
Db	61	CCAGAGCGGGAGGTAGTGGTTCCTCCGACTCGGAGCTCAGGACTCATTTTTCAGATCAGACATTT	120	
Qy	121	TACTGGGGGGTCCCGAGGACTCCGGGGATCAGGACTCATTTTTCAGATCAGACATTT	180	
Db	121	TACTGGGGGGTCCCGAGGACTCCGGGGATCAGGACTCATTTTTCAGATCAGACATTT	180	
Qy	181	CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTCTTGGCTCCGCTTCTCC	240	
Db	181	CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTCTTGGCTCCGCTTCTCC	240	
Qy	241	ACTGAGCATCTGGCGTCCCGCTCCCGCTCCAGGGGCTCACCGGGGGCTCTTCAGI	300	

Db 241 ACTGACATCTGGGCGTCCCTCCAGGGGCTCACGGGGCTTCCAGACCTGCGAGCG 300
 QY TGGCTTCTATTCTGGGACGTGAACCGGAGCCGAGCTCGTTCGCTGCTGTGAGCCTGTGC 360
 Db 301 TGGCTTCTATTCTGGGACGTGAACCGGAGCCGAGCTCGTTCGCTGCTGTGAGCCTGTGC 360
 QY GGGGGCTCCGGGAGCCTTTGGCTACCGAGCGCGAGTAGTATGTCATTAGCCCGCTGCC 420
 Db 361 GGGGGCTCCGGGAGCCTTTGGCTACCGAGCGCGAGTAGTATGTCATTAGCCCGCTGCC 420
 QY AATGCTAGCGCGCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 480
 Db 421 AATGCTAGCGCGCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 480
 QY GGTGTTCCGGGGGGGCTTCCGGAGACCCACCTCTCGCTGGGGGTGGCTTCGGGCTGG 540
 Db 481 GGTGTTCCGGGGGGGCTTCCGGAGACCCACCTCTCGCTGGGGGTGGCTTCGGGCTGG 540
 QY AACCCCGCCATCCTACGGGCGCTTGAACCCCTTACAGCGCGCGCGCGCGCGCTTCGGGAG 600
 Db 541 AACCCCGCCATCCTACGGGCGCTTGAACCCCTTACAGCGCGCGCGCGCGCGCTTCGGGAG 600
 QY AGTCGTAGCGCGCGAGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 Db 601 AGTCGTAGCGCGCGAGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 QY GAGACGCTGGTGTGCGGACGAGTCAATGGTCAAGTTCCAGCGCGCGCGCGCGCGCGCG 720
 Db 661 GAGACGCTGGTGTGCGGACGAGTCAATGGTCAAGTTCCAGCGCGCGCGCGCGCGCGCG 720
 QY TATCTGCTGACGCTGCTGCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
 Db 721 TATCTGCTGACGCTGCTGCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
 QY CCATCAACATGCTGTGCTGAAGTGTGCTTCTTAGAGTCTGCTGCTGCTGCTGCTGCTG 840
 Db 781 CCATCAACATGCTGTGCTGAAGTGTGCTTCTTAGAGTCTGCTGCTGCTGCTGCTGCTG 840
 QY GTCAACGCAATCGCGCCCTGACGCTGCGCAACTTCTGTGCTGCGCAAGAGCTGAAC 900
 Db 841 GTCAACGCAATCGCGCCCTGACGCTGCGCAACTTCTGTGCTGCGCAAGAGCTGAAC 900
 QY AAGTGAGTGACAGCAACCGCGAGTCTGGGACATGCGCATCTCTTCCAGGCGAGGAC 960
 Db 901 AAGTGAGTGACAGCAACCGCGAGTCTGGGACATGCGCATCTCTTCCAGGCGAGGAC 960
 QY CTGTGTGGAGCCACCATCTGTACACCTGGGCATGGCTGATGTGGGTACCATGTGTGAC 1020
 Db 961 CTGTGTGGAGCCACCATCTGTACACCTGGGCATGGCTGATGTGGGTACCATGTGTGAC 1020
 QY CCAAGAGAAGCTGCTCTGCTCATGAGGACGATGGCTTCCATCAGCCCTTCCAGCTGCC 1080
 Db 1021 CCAAGAGAAGCTGCTCTGCTCATGAGGACGATGGCTTCCATCAGCCCTTCCAGCTGCC 1080
 QY CAGAGCTGGGCGACGTGTTCAACATGCCCCATGACATGTAAGTGTGAGGAGGTG 1140
 Db 1081 CAGAGCTGGGCGACGTGTTCAACATGCCCCATGACATGTAAGTGTGAGGAGGTG 1140
 QY TTTGGGAAGCTCCGAGCCCAACACATGATGCCGACCTCATCCAGATCGACCGTGC 1200
 Db 1141 TTTGGGAAGCTCCGAGCCCAACACATGATGCCGACCTCATCCAGATCGACCGTGC 1200
 QY AACCCCTGGTCAAGCTGCAAGTGTGCTCATCAGCGCTTCCCTGGAGAGCGGCGCGGT 1260
 Db 1201 AACCCCTGGTCAAGCTGCAAGTGTGCTCATCAGCGCTTCCCTGGAGAGCGGCGCGGT 1260
 QY GACTGCTCTGAGCAACCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 1320
 Db 1261 GACTGCTCTGAGCAACCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 1320
 QY AGCTACACCTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 1380

Db 1321 AGCTACACCTGAGCCAGCAGTGGGAGCTGGCTTTTGGCGTGGGCTCCAAAGCCCTGTCT 1380
 QY TACATGCACTACTGACCAAGAGCTGTGTGACCGGGAAGCCCAAGGACATATGCTGTGC 1440
 Db 1381 TACATGCACTACTGACCAAGAGCTGTGTGACCGGGAAGCCCAAGGACATATGCTGTGC 1440
 QY CAGACCCGCGCACTTCCCTGGGCGGATGGCAACAGCTGTGGCGAGGGCAAGCTTGGCTC 1500
 Db 1441 CAGACCCGCGCACTTCCCTGGGCGGATGGCAACAGCTGTGGCGAGGGCAAGCTTGGCTC 1500
 QY AAAGGGGCGCTGCGTGGAGAGACACAACCTCAACAAGCACAGGGTGGATGCTTCTGGGCC 1560
 Db 1501 AAAGGGGCGCTGCGTGGAGAGACACAACCTCAACAAGCACAGGGTGGATGCTTCTGGGCC 1560
 QY AAATGGGATCCCTATGCGCCCTGCTCGCACATGTGCTGGGGCGCTGCAAGCTGGCCAGG 1620
 Db 1561 AAATGGGATCCCTATGCGCCCTGCTCGCACATGTGCTGGGGCGCTGCAAGCTGGCCAGG 1620
 QY AGCAGTGCAACCAACCCCTGCGCAAGGGGGCAAGTACTGCGAGGAGGTGAGGGTG 1680
 Db 1621 AGCAGTGCAACCAACCCCTGCGCAAGGGGGCAAGTACTGCGAGGAGGTGAGGGTG 1680
 QY AAATACCGATCCCTGCAACCTGGAGCCCTGCCAGCTCAGCTTCCGGAAAGAGCTTCCGG 1740
 Db 1681 AAATACCGATCCCTGCAACCTGGAGCCCTGCCAGCTCAGCTTCCGGAAAGAGCTTCCGG 1740
 QY GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCACCACCGGCTCACTCTCGCC 1800
 Db 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCACCACCGGCTCACTCTCGCC 1800
 QY GTGCAATGGTGCCCAAGTACTCGGGGCTCTCCCGGAGCAAGTGCAGCTCACTGTGC 1860
 Db 1801 GTGCAATGGTGCCCAAGTACTCGGGGCTCTCCCGGAGCAAGTGCAGCTCACTGTGC 1860
 QY CGACCAATGCACTGCTTCTATGTGCTGCAAGCTTCCAGGCTGGGAGCGCACGCTG 1920
 Db 1861 CGACCAATGCACTGCTTCTATGTGCTGCAAGCTTCCAGGCTGGGAGCGCACGCTG 1920
 QY TGCTCTCTGACTCCACTCGCTGCTCAAGGCAAGTGCATCAAGGCTGGCTGTGAT 1980
 Db 1921 TGCTCTCTGACTCCACTCGCTGCTCAAGGCAAGTGCATCAAGGCTGGCTGTGAT 1980
 QY GGAACCTGGGCTCCAGAGAGATTCGCAAGTGTGGGGTGTGGGGGAGACAATAAG 2040
 Db 1981 GGAACCTGGGCTCCAGAGAGATTCGCAAGTGTGGGGTGTGGGGGAGACAATAAG 2040
 QY AGCTGCAAGAGGTGACTGGAATCTTCAAGGCCATGATGGCTACAAATTTCTGTTG 2100
 Db 2041 AGCTGCAAGAGGTGACTGGAATCTTCAAGGCCATGATGGCTACAAATTTCTGTTG 2100
 QY GCAATCCCGCAGCGCTCAAGCATCGACATCGCGAGCGGTTTCAAAAGGCGCTGATC 2160
 Db 2101 GCAATCCCGCAGCGCTCAAGCATCGACATCGCGAGCGGTTTCAAAAGGCGCTGATC 2160
 QY GGGATGACAACTACTGGCTTGAAGAACAGCAAGCAAGTACCTGCTCAAGGGGAT 2220
 Db 2161 GGGATGACAACTACTGGCTTGAAGAACAGCAAGCAAGTACCTGCTCAAGGGGAT 2220
 QY TTCTGTGTGTGCGGGTGGAGCGGACCTGGTGTGAAGGGCAGTCTGCTGCGGTACAGC 2280
 Db 2221 TTCTGTGTGTGCGGGTGGAGCGGACCTGGTGTGAAGGGCAGTCTGCTGCGGTACAGC 2280
 QY GGACAGGGCAGCGGTGGAGAGCTTCAGGCTTTCGGGCGGCAATCTTGGAGCGCGTGACC 2340
 Db 2281 GGACAGGGCAGCGGTGGAGAGCTTCAGGCTTTCGGGCGGCAATCTTGGAGCGCGTGACC 2340
 QY GTGAGGTCTCTCCGTGGGAGAGTACACCGCCCGGGTCCGCTACTCTTCTATCTG 2400
 Db 2341 GTGAGGTCTCTCCGTGGGAGAGTACACCGCCCGGGTCCGCTACTCTTCTATCTG 2400
 QY CCAAAAGAGCTCCGGGAGGACAAAGTCTCTATCCCAAGGACCCCGGGGACCTCTGTG 2460
 Db 2401 CCAAAAGAGCTCCGGGAGGACAAAGTCTCTATCCCAAGGACCCCGGGGACCTCTGTG 2460

```
QY 2461 TTGCACAAACAGCTGCTCAGCTCTCCAAACAGGTGGACAGCCGACGACAGGCCCTT 2520
|
Db 2461 TTGCACAAACAGCTGCTCAGCTCTCCAAACAGGTGGACAGCCGACGACAGGCCCTT 2520
|
QY 2521 GCACGCTGGTGGCTGGCAGCTGGGGCGGTGTCTCCGCGAGCTGCGGCGAGTGGCTTGCAG 2580
|
Db 2521 GCACGCTGGTGGCTGGCAGCTGGGGCGGTGTCTCCGCGAGCTGCGGCGAGTGGCTTGCAG 2580
|
QY 2581 AAGCGGCGGTGACGTCCGGGGCTCCGCGGCGACGACAGGTGCTGCTGCTGATGCA 2640
|
Db 2581 AAGCGGCGGTGACGTCCGGGGCTCCGCGGCGACGACAGGTGCTGCTGCTGATGCA 2640
|
QY 2641 GCCCATCGGCGGTGGAGACACAAAGCTGCGGGGAGCCCTGCCCACTGGGAGCTCAGC 2700
|
Db 2641 GCCCATCGGCGGTGGAGACACAAAGCTGCGGGGAGCCCTGCCCACTGGGAGCTCAGC 2700
|
QY 2701 GCCTGGTCAACCTGCTCCAAAGAGCTGGGGCGGGGATTCAGAGGGCTCACTCAAGTGT 2760
|
Db 2701 GCCTGGTCAACCTGCTCCAAAGAGCTGGGGCGGGGATTCAGAGGGCTCACTCAAGTGT 2760
|
QY 2761 GTGGGCGACGAGCGCGGTGCTGGCGCGGACAGTGAAGTTCAGAGGGCTCACTCAAGTGT 2820
|
Db 2761 GTGGGCGACGAGCGCGGTGCTGGCGCGGACAGTGAAGTTCAGAGGGCTCACTCAAGTGT 2820
|
QY 2821 GAGCTGGACTTTCGCTGCTGAGGCGGTGCTGA 2853
|
Db 2821 GAGCTGGACTTTCGCTGCTGAGGCGGTGCTGA 2853
|
```

RESULT 2

```
US-09-965-631-7
; Sequence 7, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Encod
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 3446
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-965-631-7
```

```
Query Match 99.8%; Score 2846.6; DB 10; Length 3446;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2849; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 ATGCTTTTCTGGGATCTTAACCTGGCTTTTCGGCGGCGAAGCGTGGAGGCTTTGAG 60
|
Db 397 ATGCTTTCTGCTGGGATCTTAACCTGGCTTTTCGGCGGCGAAGCGTGGAGGCTCTGAG 456
|
QY 61 CCAGAGCGGAGGTAGTCTCCATCGACCTGGACCGGACATTAAGCGCGCGCTTAC 120
|
Db 457 CCAGAGCGGAGGTAGTCTCCATCGACCTGGACCGGACATTAAGCGCGCGCTTAC 516
|
QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTATTTTCAGATCACAGCATTT 180
|
Db 517 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTATTTTCAGATCACAGCATTT 576
|
QY 181 CAGGAGGACTTTTACTACACTGACGCGGATGCTAGTCTTGGCTCCCGGCTTCTCC 240
|
Db 577 CAGGAGGACTTTTACTACACTGACGCGGATGCTAGTCTTGGCTCCCGGCTTCTCC 636
|
QY 241 ACTGAGCATCTGGGCTTCCCTCCAGGGCTTCACGGGGCTCTTCAGACCTGGAGCG 300
|
```

```
Db 637 ACTGAGCATCTGGGCTTCCCTCCAGGGGCTCACCAGGGGCTCTTTCAGTCTCCAGGCG 696
|
QY 301 TGCCTTCTATTCTGGGAGCTGAACCGCGAGCGGACTGTTTCGGTGTGTGAGCTGTGCG 360
|
Db 697 TGCCTTCTATTCTGGGAGCTGAACCGCGAGCGGACTGTTTCGGTGTGTGAGCTGTGCG 756
|
QY 361 GGGGGCTCCGCGGAGCTTTGGCTACCGAGCGCGGAGTATGTCATTAGCCCGCTGCC 420
|
Db 757 GGGGGCTCCGCGGAGCTTTGGCTACCGAGCGCGGAGTATGTCATTAGCCCGCTGCC 816
|
QY 421 AATGCTAGCGGCGCGCGCGCAGCGCAACAGCCAGGCGCACACCTTCTCCAGCGCGG 480
|
Db 817 AATGCTAGCGCGCGCGCGCAGCGCAACAGCCAGGCGCACACCTTCTCCAGCGCGG 876
|
QY 481 GGTGTTCCGGGCGGCTTCCGGAGACCCACCTCTCCTCGGGGTGGGCTCGGGTGG 540
|
Db 877 GGTGTTCCGGGCGGCGCTTCCGGAGACCCACCTCTCCTCGGGGTGGGCTCGGGTGG 936
|
QY 541 AACCCCGCATCTTACGGGCGCTTGACCCCTTACAAGCGCGCGGCGGGGCTTCCGGGAG 600
|
Db 937 AACCCCGCATCTTACGGGCGCTTGACCCCTTACAAGCGCGCGGCGGGGCTTCCGGGAG 996
|
QY 601 AGTCTAGCGCGCGAGCTTGGGCGCGCAAGCTTTTCGTGTCTATCCCGCGTACGTG 660
|
Db 997 AGTCTAGCGCGCGAGCTTGGGCGCGCAAGCTTTTCGTGTCTATCCCGCGTACGTG 1056
|
QY 661 GAGAGCTGGTGGTCCGGGACGAGTCAATGTCAAGTTCACGCGCGGCGGACCTGGACAT 720
|
Db 1057 GAGAGCTGGTGGTCCGGGACGAGTCAATGTCAAGTTCACGCGCGGCGGACCTGGACAT 1116
|
QY 721 TATCTGCTAGCGCTGCTGGCAACGCGCGGCGGACTCTACCGCCATCCAGCATCTCTCAAC 780
|
Db 1117 TATCTGCTAGCGCTGCTGGCAACGCGCGGCGGACTCTACCGCCATCCAGCATCTCTCAAC 1176
|
QY 781 CCCATCAACATCTGTTGGTCAAGTGTGCTTCTTAGAGATCTGACTCCGGGCGCAAG 840
|
Db 1177 CCCATCAACATCTGTTGGTCAAGTGTGCTTCTTAGAGATCTGACTCCGGGCGCAAG 1236
|
QY 841 GTCACCGCATGGGCGCTCAGCTGGCAACTTCTGTGCTGGCAGAGAACTGAAC 900
|
Db 1237 GTCACCGCATGGGCGCTCAGCTGGCAACTTCTGTGCTGGCAGAGAACTGAAC 1296
|
QY 901 AAAGTGAAGTGAAGACCCCGAGTACTGGACACTGCCATCTCTTCAACAGGACGAG 960
|
Db 1297 AAAGTGAAGTGAAGACCCCGAGTACTGGGACACTGCCATCTCTTCAACAGGACGAG 1356
|
QY 961 CTGTGTGGAGCCACCTGTGACACCTGGGATGGCTGATGGGTACATGTGTGAC 1020
|
Db 1357 CTGTGTGGAGCCACCTGTGACACCTGGGATGGCTGATGGGTACATGTGTGAC 1416
|
QY 1021 CCCAAGAGAAGCTGCTGTCTATTGAGAGATGGGCTTCCATCAGCTTCACTGTC 1080
|
Db 1417 CCCAAGAGAAGCTGCTGTCTATTGAGAGATGGGCTTCCATCAGCTTCACTGTC 1476
|
QY 1081 CACGAGCTGGGCCAGCTGTTCAACATGCCCATGCCATGACAATGTGAAAGTCTGTGAGAGGTG 1140
|
Db 1477 CACGAGCTGGGCCAGCTGTTCAACATGCCCATGCCATGACAATGTGAAAGTCTGTGAGAGGTG 1536
|
QY 1141 TTTGGGAGCTCCGAGCAACCATCATGTCCCGGACCTTCATCCAGATCGACCGTGC 1200
|
Db 1537 TTTGGGAGCTCCGAGCAACCATCATGTCCCGGACCTTCATCCAGATCGACCGTGC 1596
|
QY 1201 AACCCCTGGTCCAGCTGCTGAGTGTCCATCATACCCACTTCTCTGGACAGCGGACGCT 1260
|
Db 1597 AACCCCTGGTCCAGCTGCTGAGTGTCCATCATACCCACTTCTCTGGACAGCGGACGCT 1556
|
QY 1261 GACTGCTCTCTGGACCAACCGAGCAAGCCCTCTCTCCCGGAGGATCTGCGGGGCGCC 1320
|
Db 1657 GACTGCTCTCTGGACCAACCGAGCAAGCCCTCTCTCCCGGAGGATCTGCGGGGCGCC 1716
|
QY 1321 AGCTACACCTGAGCCAGCAGTGGAGCTGGCTTTTGGCGTGGGCTCCAAAGCTCTGCT 1380
|
Db 1717 AGCTACACCTGAGCCAGCAGTGGAGCTGGCTTTTGGCGTGGGCTCCAAAGCTCTGCT 1776
|
```

QY 1381 TACATGACGACTGACCAAGCTGTGTGACCGGGAAGGCCAAGGACAGATGTTGTGC 1440
 DB 1777 TACATGACGACTGACCAAGCTGTGTGACCGGGAAGGCCAAGGACAGATGTTGTGC 1836
 QY 1441 CAGACCGCCGACATTCCTCCCTGGGCGGATGACCAAGCTGTGGGAGGACAGTCTGCCCTC 1500
 DB 1837 CAGACCGCCGACATTCCTCCCTGGGCGGATGACCAAGCTGTGGGAGGACAGTCTGCCCTC 1896
 QY 1501 AAGGGGCCCTGCTGGAGAGACACACCTCAACAGCAGAGGTGATGTTCTCTGGGCC 1560
 DB 1897 AAGGGGCCCTGCTGGAGAGACACACCTCAACAGCAGAGGTGATGTTCTCTGGGCC 1956
 QY 1561 AATATGGATCCCTATATGCCCCCTGCTCGCGCACATGTGTGGGGCGCTGACGTGGCCAGG 1620
 DB 1957 AATATGGATCCCTATATGCCCCCTGCTCGCGCACATGTGTGGGGCGCTGACGTGGCCAGG 2016
 QY 1621 AGCGATGTCACCAACCCACCCCTGCCAAGCGGGGCAAGTACTGCGAGGAGTGAAGGTG 1680
 DB 2017 AGCGATGTCACCAACCCACCCCTGCCAAGCGGGGCAAGTACTGCGAGGAGTGAAGGTG 2076
 QY 1681 AATATCCGATCCCTGACACCTGAGCCCTGCCCCAGCTCAGCCTCCGAAAGAGCTTCCGG 1740
 DB 2077 AATATCCGATCCCTGACACCTGAGCCCTGCCCCAGCTCAGCCTCCGAAAGAGCTTCCGG 2136
 QY 1741 GAGGACGAGTGTGAGGCTTTCAACGGCTACACCAAGCAGCAGCAACCGCGCTCACTCTCGCC 1800
 DB 2137 GAGGACGAGTGTGAGGCTTTCAACGGCTACACCAAGCAGCAGCAACCGCGCTCACTCTCGCC 2196
 QY 1801 GTGCAATGGGTGCCAAGTACTCGGGGTGTCTCCCGGGGACAAGTGCAAGCTCATCTGC 1860
 DB 2197 GTGCAATGGGTGCCAAGTACTCGGGGTGTCTCCCGGGGACAAGTGCAAGCTCATCTGC 2256
 QY 1861 CGACCAATGGCACTGCTACTTCTATGTCTGTGCAACCAAGTGTGGAGCGACGCTG 1920
 DB 2257 CGACCAATGGCACTGCTACTTCTATGTCTGTGCAACCAAGTGTGGAGCGACGCTG 2316
 QY 1921 TGCTCTCTGACTCCACCTCCGCTGTGTGTCCAAAGGACAGTGCATCAAGGCTGCTGTGAT 1980
 DB 2317 TGCTCTCTGACTCCACCTCCGCTGTGTGTCCAAAGGACAGTGCATCAAGGCTGCTGTGAT 2376
 QY 1981 GGGAACTGGGCTCCAGAGAGATTCGACAGTGTGGGGTGTGGGGGAGACAATAAG 2040
 DB 2377 GGGAACTGGGCTCCAGAGAGATTCGACAGTGTGGGGTGTGGGGGAGACAATAAG 2436
 QY 2041 AGCTGCAAGAGTGTGACTGGACTCTTCCCAAGCCCATGCTGCTACATTTCTGTTG 2100
 DB 2437 AGCTGCAAGAGTGTGACTGGACTCTTCCCAAGCCCATGCTGCTACATTTCTGTTG 2496
 QY 2101 GCCATCCCGGAGCGCTCAAGCATCGACATCGCCAGCGCGGTTACAAGGGCTGATC 2160
 DB 2497 GCCATCCCGGAGCGCTCAAGCATCGACATCGCCAGCGCGGTTACAAGGGCTGATC 2556
 QY 2161 GGGATGACAACTACCTGGCTCTGAAAGACAGCAAGGCAAGTACCTGCTCAACGGGCAT 2220
 DB 2557 GGGATGACAACTACCTGGCTCTGAAAGACAGCAAGGCAAGTACCTGCTCAACGGGCAT 2616
 QY 2221 TTCGTGTGTGCGCGGTGGAGCGGACCTGCTGTTGAAGGCGAGTGTGCTGCGGTACAGC 2280
 DB 2617 TTCGTGTGTGCGCGGTGGAGCGGACCTGCTGTTGAAGGCGAGTGTGCTGCGGTACAGC 2676
 QY 2281 GGCACGGGCACAGGGTGGAGAGCTCGAGGCTTCCCGGCCCATCTCGGAGCGCTGACC 2340
 DB 2677 GGCACGGGCACAGGGTGGAGAGCTCGAGGCTTCCCGGCCCATCTCGGAGCGCTGACC 2736
 QY 2341 GTGAGGTCTCTCTCGGTGGGAAGTACACCGCGCGGGTCCGCTACTCTCTATCTG 2400
 DB 2737 GTGAGGTCTCTCTCGGTGGGAAGTACACCGCGCGGGTCCGCTACTCTCTATCTG 2796
 QY 2401 CCCAAAGAGCTCGGGAGGACAAGTCTCTCATATCCCAAGGACCCCGGGGACCCCTCTGTC 2460
 DB 2797 CCCAAAGAGCTCGGGAGGACAAGTCTCTCATATCCCAAGGACCCCGGGGACCCCTCTGTC 2856

QY 2461 TTGCACAAACGCGTCTCAGCCTCTCCAAACAGGTGGAGCAGCGGAGCAGAGGCCCCCT 2520
 DB 2857 TTGCACAAACGCGTCTCAGCCTCTCCAAACAGGTGGAGCAGCGGAGCAGAGGCCCCCT 2916
 QY 2521 GCACGCTGGGTGGAGCTGGGGCGCTGCTCCGCGAGCTGCGGAGTGGCGCTGTCAG 2580
 DB 2917 GCACGCTGGGTGGAGCTGGGGCGCTGCTCCGCGAGCTGCGGAGTGGCGCTGTCAG 2976
 QY 2581 AAGCGGCGGTGGAGCTGCGGGGCTCCCGCGGAGCAGCGCAGCTGCTGCTGTGATGCA 2640
 DB 2977 AAGCGGCGGTGGAGCTGCTGGGGCTCCCGCGGAGCAGCGCAGCTGCTGTGATGCA 3036
 QY 2641 GCCATGCGGCGGTGGAGCAGCAAGCTGCGGGGAGCCCTGCCCACTGGGAGCTCAGC 2700
 DB 3037 GCCATGCGGCGGTGGAGCAGCAAGCTGCGGGGAGCCCTGCCCACTGGGAGCTCAGC 3096
 QY 2701 GCCTGGTCACTGCTTCCAAAGCTGCGGGCGGAGTTCAGAGCGCTCACTCAAGTGT 2760
 DB 3097 GCCTGGTCACTGCTTCCAAAGCTGCGGGCGGAGTTCAGAGCGCTCACTCAAGTGT 3156
 QY 2761 GTGGGCCACGAGCGCGCTGCTGGCCCGGAGCAGTGCACCTTGCACCGCAAGCCCAAG 2820
 DB 3157 GTGGGCCACGAGCGCGCTGCTGGCCCGGAGCAGTGCACCTTGCACCGCAAGCCCAAG 3216
 QY 2821 GAGCTGGAGCTTCTGCGTCTGAGGCCGTGCTGA 2853
 DB 3217 GAGCTGGAGCTTCTGCGTCTGAGGCCGTGCTGA 3249

RESULT 3
 US-10-163-316-3
 ; Sequence 3, Application US/10163316
 ; Publication No. US2002019703A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kapeller-Libermann, Rosana
 ; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
 ; TITLE OF INVENTION: Therefor
 ; FILE REFERENCE: MP101-025P1RNM
 ; CURRENT APPLICATION NUMBER: US/10/163,316
 ; CURRENT FILING DATE: 2002-06-05
 ; PRIOR APPLICATION NUMBER: 60/297,863
 ; PRIOR FILING DATE: 2001-06-13
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 2469
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(2469)
 US-10-163-316-3

Query Match 80.5%; Score 2296.4; DB 9; Length 2469;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 2379; Conservative 0; Mismatches 6; Indels 69; Gaps 1;

QY 1 ATGCTTTTCTGGGATCTCAACCTTTCGCGGCGGAGCAGTTCGAGGCTTTGAG 60
 DB 1 ATGCTTTTCTGGGATCTCAACCTTTCGCGGCGGAGCAGTTCGAGGCTTTGAG 60
 QY 61 CCAGAGCGGAGGTAGTCTGCTCCATCCGACTGGACCGGACATTAAGCGCGCGCTAC 120
 DB 61 CCAGAGCGGAGGTAGTCTGCTCCATCCGACTGGACCGGACATTAAGCGCGCGCTAC 120
 QY 121 TACTGGCGGGTCCCGAGGACTCCCGGGATCAGGAGTCAATTTTTCAGATCAGCATTT 180
 DB 121 TACTGGCGGGTCCCGAGGACTCCCGGGATCAGGAGTCAATTTTTCAGATCAGCATTT 180
 QY 181 CAGGAGGACTTTTACCTACACCTGACCGGATGCTCAGTCTTGGCTCCCGCTTCTCC 240
 DB 181 CAGGAGGACTTTTACCTACACCTGACCGGATGCTCAGTCTTGGCTCCCGCTTCTCC 240

QY 241 ACTGAGCATCTGGCGCTCCCTCCAGGGCTCACCGGGCTCTTCAGACCTCGCAGCG 300
DB 241 ACTGAGCATCTGGCGCTCCCTCCAGGGCTCACCGGGCTCTTCAGACCTCGCAGCG 300
QY 301 TGCCTCTATTCTGGGACGTGAACCGCGAGCGGACTCGTTCGCTGTGAGCCTGTGC 360
DB 301 TGCCTCTATTCTGGGACGTGAACCGCGAGCGGACTCGTTCGCTGTGAGCCTGTGC 360
QY 361 GGGGGCTCCGCGAGCCTTTGGCTACCGAGGGCGCGAGTATGTCTATTAGCCGCTGCC 420
DB 361 GGGGGCTCCGCGAGCCTTTGGCTACCGAGGGCGCGAGTATGTCTATTAGCCGCTGCC 420
QY 421 AATGCTAGCGCGCGCGGCGCGAGCGCAACAGCCAGGGCGCACACTTCTCCAGCGCG 480
DB 421 AATGCTAGCGCGCGCGGCGCGAGCGCAACAGCCAGGGCGCACACTTCTCCAGCGCG 480
QY 481 GGTGTTCCGGCGCGGCTTCGCGAGACCCACCTCTCGCTGCGGGGTGGCTCGGGCTGG 540
DB 481 GGTGTTCCGGCGCGGCTTCGCGAGACCCACCTCTCGCTGCGGGGTGGCTCGGGCTGG 540
QY 541 AACCCGCCATCTACGGGCTCGGACCTTTACAAGCGCGCGGGCTTCGCGGAG 600
DB 541 AACCCGCCATCTACGGGCTCGGACCTTTACAAGCGCGCGGGCTTCGCGGAG 600
QY 601 AGTCGTAGCGCGCGAGGCTCGGGCGCGCAAGCGTTTGTGTCTATCCGCGGTACGTG 660
DB 601 AGTCGTAGCGCGCGAGGCTCGGGCGCGCAAGCGTTTGTGTCTATCCGCGGTACGTG 660
QY 661 GAGACGCTGTGTGCGGGACGAGTCAATGGTCAAGTTCACCGCGCGGACCTGGAAAT 720
DB 661 GAGACGCTGTGTGCGGGACGAGTCAATGGTCAAGTTCACCGCGCGGACCTGGAAAT 720
QY 721 TATCTGCTGACGCTGTGGCAACGGCGGCGGACTCTACCGCATCCAGCATCTCTCAAC 780
DB 721 TATCTGCTGACGCTGTGGCAACGGCGGCGGACTCTACCGCATCCAGCATCTCTCAAC 780
QY 781 CCCATCAACATCTGTGTGTCAGGTGCTGCTCTTAGAGATCGTGACTCCGGGCGCAAG 840
DB 781 CCCATCAACATCTGTGTGTCAGGTGCTGCTCTTAGAGATCGTGACTCCGGGCGCAAG 840
QY 841 GTCACGGCAATCGGCGCTGACGCTGCGCACTTCTGTGCTGGCAGAAAGCTGAAC 900
DB 841 GTCACGGCAATCGGCGCTGACGCTGCGCACTTCTGTGCTGGCAGAAAGCTGAAC 900
QY 901 AAAGTGAGTGAAGCAACCCGAGTACTGGGACACTGCCATCCTCTTACCAGGCGAGAC 960
DB 901 AAAGTGAGTGAAGCAACCCGAGTACTGGGACACTGCCATCCTCTTACCAGGCGAGAC 960
QY 961 CTGTGTGAGGCCACACCTGTGACACCTGGGCATGGCTGATGGGTACCATGTGTGAC 1020
DB 961 CTGTGTGAGGCCACACCTGTGACACCTGGGCATGGCTGATGGGTACCATGTGTGAC 1020
QY 1021 CCCAAGAGAGCTGCTCTGTATTGAGGACGATGGGCTTCCATCAGCCTTACCACCTGCC 1080
DB 1021 CCCAAGAGAGCTGCTCTGTATTGAGGACGATGGGCTTCCATCAGCCTTACCACCTGCC 1080
QY 1081 CAGAGCTGGGCGACGTTTCAACATGCCCATGACAAATGTGAAGTCTGTGAGGAGTG 1140
DB 1081 CAGAGCTGGGCGACGTTTCAACATGCCCATGACAAATGTGAAGTCTGTGAGGAGTG 1140
QY 1141 TTTGGGAAGCTCCGAGCCACACATGATGTCGCCAGCTTCCATCAGCCTTACCACCTGCC 1200
DB 1141 TTTGGGAAGCTCCGAGCCACACATGATGTCGCCAGCTTCCATCAGCCTTACCACCTGCC 1200
QY 1201 AACCCCTGTGAGCTGAGTGTGCTGCCATCATCAGGACTTCTTGGACAGCGGCGACGGT 1260
DB 1201 AACCCCTGTGAGCTGAGTGTGCTGCCATCATCAGGACTTCTTGGACAGCGGCGACGGT 1260
QY 1261 GACTGCTCTTGAGCAACCCAGCAAGCCCATCTCCCTGCGGAGGATCTGCGGGCGCC 1320
DB 1261 GACTGCTCTTGAGCAACCCAGCAAGCCCATCTCCCTGCGGAGGATCTGCGGGCGCC 1320
QY 1321 AGCTACACCTTGAGCGAGGCTGGGCTTTTGGCGGTGCCAGCCCTGTCTCT 1380

DB 1321 AGCTACACCTTGAGCGAGGCTGGGCTTTTGGCGGTGCCAGCCCTGTCTCT 1380
QY 1381 TACATGCACTACTGACCAAGCTGTGTGCACCGGGAAGGCCAAGGAGACAGATGTGTGC 1440
DB 1381 TACATGCACTACTGACCAAGCTGTGTGCACCGGGAAGGCCAAGGAGACAGATGTGTGC 1440
QY 1441 CAGACCGCGCACTTCCCTGGGCGGATGGCACCACTGTGGGAGGGGAAGCTGTGCCTC 1500
DB 1441 CAGACCGCGCACTTCCCTGGGCGGATGGCACCACTGTGGGAGGGGAAGCTGTGCCTC 1500
QY 1501 AAAGGGCGCTGGTGGAGAGACACAACCTTCAACAAGCACAG----- 1541
DB 1501 AAAGGGCGCTGGTGGAGAGACACAACCTTCAACAAGCACAG----- 1541
QY 1542 -----GTTGGATGGT 1551
DB 1561 TCTCCAAAACAACCTTATTAAAGGCTACCAATGGCTGCACACTTACACAGGTGGATGGT 1620
QY 1562 TCTTGGGCAATGGATCCCTATGCGCCCTGCTGCGGCACATGTGGTGGGGGCTGCAG 1611
DB 1621 TCTTGGGCAATGGATCCCTATGCGCCCTGCTGCGGCACATGTGGTGGGGGCTGCAG 1680
QY 1612 CTGGCCAGGAGGCTGACCAACCCACCTCTGCCAAGGGGCAAGTACTGCGAGGGA 1671
DB 1681 CTGGCCAGGAGGCTGACCAACCCACCTCTGCCAAGGGGCAAGTACTGCGAGGGA 1740
QY 1672 GTGAGGGTGAATACCGATCTGCAACCTGGAGCCCTGCCAGCTAGCTCCGGGAAG 1731
DB 1741 GTGAGGGTGAATACCGATCTGCAACCTGGAGCCCTGCCAGCTAGCTCCGGGAAG 1800
QY 1732 AGCTTCCGGGAGGAGCTGTGAGGCTTCAACGGCTTCAACACGACAGCAACCAAGGCTC 1791
DB 1801 AGCTTCCGGGAGGAGCTGTGAGGCTTCAACGGCTTCAACACGACAGCAACCAAGGCTC 1860
QY 1792 ACTCTCGCGTGGCATGGGTGCCAAGTACTCCGGCGCTGCTCCCGGGGCAAGTGAAG 1851
DB 1861 ACTCTCGCGTGGCATGGGTGCCAAGTACTCCGGCGCTGCTCCCGGGGCAAGTGAAG 1920
QY 1852 CTCATCTCCGAGCAATGGCAGCTGGCTACTTCTATGTGTGGACACCAAGGCTGGAC 1911
DB 1921 CTCATCTCCGAGCAATGGCAGCTGGCTACTTCTATGTGTGGACACCAAGGCTGGAC 1980
QY 1912 GGCAGCTGTGCTCTCTGACTCCACCTCCGCTGTGTGTCACCAAGCAAGTGCATCAAGCT 1971
DB 1981 GGCAGCTGTGCTCTCTGACTCCACCTCCGCTGTGTGTCACCAAGCAAGTGCATCAAGCT 2040
QY 1972 GCCTGTGATGGGAACCTGGGCTCCAAAGAGATTCGACAAAGTGTGGGCTGTG*GGGGGA 2031
DB 2041 GCCTGTGATGGGAACCTGGGCTCCAAAGAGATTCGACAAAGTGTGGGCTGTG*GGGGGA 2100
QY 2032 GACAAATAGAGCTGCAAGAAGTGAAGTGTGCTTCCAGGCAAGTGCATGCTACAAAT 2091
DB 2101 GACAAATAGAGCTGCAAGAAGTGAAGTGTGCTTCCAGGCAAGTGCATGCTACAAAT 2160
QY 2092 TTCGTGTGGCCATCCCGCAGGCGCTCAAGCATCGACATCCCGCAGCGGTTACAAA 2151
DB 2161 TTCGTGTGGCCATCCCGCAGGCGCTCAAGCATCGACATCCCGCAGCGGTTACAAA 2220
QY 2152 GGGCTGATCGGGGATGACAACTACTGGCTCTGAAGAACAGCAAGTACCTGCTC 2211
DB 2221 GGGCTGATCGGGGATGACAACTACTGGCTCTGAAGAACAGCAAGTACCTGCTC 2280
QY 2212 AACGGGCAATTCGTGGTGTGGCGGCTGGAGCGGCACTGTGGTGAAGGCAAGTGTGCTG 2271
DB 2281 AACGGGCAATTCGTGGTGTGGCGGCTGGAGCGGCACTGTGGTGAAGGCAAGTGTGCTG 2340
QY 2272 CGGTACAGCGGACAGCGGCTGGAGAGCTTGAAGCTTCCCGGCGGCTTCCCTGGAG 2331
DB 2341 CGGTACAGCGGACAGCGGCTGGAGAGCTTGAAGCTTCCCGGCGGCTTCCCTGGAG 2400
QY 2332 CCGCTGACCTGGAGGCTCTCTCGTGGGGAAGATGACACCGCGCGGCTCCCG 2385
DB 2385 CCGCTGACCTGGAGGCTCTCTCGTGGGGAAGATGACACCGCGCGGCTCCCG

Db 2401 CCGCTGACCGTGGAGTCTCTCCGTGGGAGAGATGACACCGCCGCGGACCTGC 2454

RESULT 4

US-10-163-316-1

; Sequence 1, Application US/10163316

; Publication No. US20020197703A1

; GENERAL INFORMATION:

; APPLICANT: Kapeller-Libermann, Rosana

; TITLE OF INVENTION: 65522, A Human Matrix Metalloproteinase and Uses

; FILE OF INVENTION: Therefor

; FILE REFERENCE: MPI01-025PIRNM

; CURRENT APPLICATION NUMBER: US/10/163,316

; CURRENT FILING DATE: 2002-06-05

; PRIOR APPLICATION NUMBER: 60/297,863

; PRIOR FILING DATE: 2001-06-13

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 2940

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (472)...(2941)

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(2940)

; OTHER INFORMATION: n = A, T, C or G

US-10-163-316-1

Query Match 80.5%; Score 2296.4; DB 9; Length 2940;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 2379; Conservative 0; Mismatches 6; Indels 69; Gaps 1;

Qy 1 ATGCTTTTGTGGCATCCTTAACCTCGCTTTCGCCGGGGAACCCGTGGAGGCTTTGAG 60
Db 472 ATGCTTTCGTGGGCATCCTTAACCTCGCTTTCGCCGGGGAACCCGTGGAGGCTTTGAG 531
Qy 61 CCAGAGGGGAGGTAGTCTTCCATCCGACTGGACCCGGGACATTAACGGCCCGCGCTAC 120
Db 532 CCAGAGGGGAGGTAGTCTTCCATCCGACTGGACCCGGGACATTAACGGCCCGCGCTAC 591
Qy 121 TACTGGGGGGTCCCGAGGACTCCGGGGATCAGGACTCATTTTTCAGATCAGACATTT 180
Db 592 TACTGGGGGGTCCCGAGGACTCCGGGGATCAGGACTCATTTTTCAGATCAGACATTT 651
Qy 181 CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTCTTTGGCTCCGCGCTTCTCC 240
Db 652 CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTCTTTGGCTCCGCGCTTCTCC 711
Qy 241 ACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGCGACGC 300
Db 712 ACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGCGACGC 771
Qy 301 TGTCTTATTCCTGGGAGCTGAACCGCGAGCGGACTCTGTCGCTGTGTGAGGCTGTGC 360
Db 772 TGTCTTATTCCTGGGAGCTGAACCGCGAGCGGACTCTGTCGCTGTGTGAGGCTGTGC 831
Qy 361 GGGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGGATGATGATTTAGCCGCTGCC 420
Db 832 GGGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGGATGATGATTTAGCCGCTGCC 891
Qy 421 AATGCTAGCGCGGGGCGAGCGAACAACCGAGGCGGCACACCTTCTCCAGCGCGG 480
Db 892 AATGCTAGCGCGGGGCGAGCGAACAACCGAGGCGGCACACCTTCTCCAGCGCGG 951
Qy 481 GGTGTTCCGGGCGGGCTTCCGAGAGACCCCACTCTCTCGCTCGGGGTGGGCTCGGGCTG 540
Db 952 GGTGTTCCGGGCGGGCTTCCGAGAGACCCCACTCTCTCGCTCGGGGTGGGCTCGGGCTG 1011
Qy 541 AACCCCGCATCTACGGGCTCGACCTTTACAAAGCGCGGCGGGGCTTCGGGGAG 600
Db 1011

Db 1012 AACCCCGCATCTCTACGGGCTTGAACCTTTACAAAGCGCGGCGGGGCTTCGGGGAG 1071
Qy 601 AGTCGTAGCGCGCAGGTCTGGCGCGCCCAAGGTTTCGTGTCTATATCCCTCGGTACGTG 660
Db 1072 AGTCGTAGCGCGCAGGTCTGGCGCGCCCAAGGTTTCGTGTCTATATCCCTCGGTACGTG 1131
Qy 661 GAGACGCTGGTGGTCCGCGAGAGTCAATGTCAAGTTTCCACGGCGCGGAGCTTGAACAT 720
Db 1132 GAGACGCTGGTGGTCCGCGAGAGTCAATGTCAAGTTTCCACGGCGCGGAGCTTGAACAT 1191
Qy 721 TATCTGCTGACGCTCTGCGACGCGGCGGACTCTACCGCATCCAGCATCTCTCAAC 780
Db 1192 TATCTGCTGACGCTCTGCGACGCGGCGGACTCTACCGCATCCAGCATCTCTCAAC 1251
Qy 781 CCCATCAACATCGTTGTGCTCAAGTGTCTTCTTAGAGATCGTGTGCTCGGGGCCCAAG 840
Db 1252 CCCATCAACATCGTTGTGCTCAAGTGTCTTCTTAGAGATCGTGTGCTCGGGGCCCAAG 1311
Qy 841 GTCACGGGAATGCGGCGCTGACGCTGGCAACTTCTGTGCTGCGCAAGAAGCTGAAC 900
Db 1312 GTCACGGGAATGCGGCGCTGACGCTGGCAACTTCTGTGCTGCGCAAGAAGCTGAAC 1371
Qy 901 AAAGTGAGTGACAAGCACCCCGAGTACTGGGACACTGCGATCTCTTCCACGAGGACGAC 960
Db 1372 AAAGTGAGTGACAAGCACCCCGAGTACTGGGACACTGCGATCTCTTCCACGAGGACGAC 1431
Qy 961 CTGTGTGGAGCCACCACTGTGACACCTTGGCATGGGTGTGCTGGGTACCATGTGTGAC 1020
Db 1432 CTGTGTGGAGCCACCACTGTGACACCTTGGCATGGGTGTGCTGGGTACCATGTGTGAC 1491
Qy 1021 CCCAAGAGAGCTGCTGCTGATTCAGGAGGATGGGCTTCCATAGCCTTCCACTGCTGCC 1080
Db 1492 CCCAAGAGAGCTGCTGCTGATTCAGGAGGATGGGCTTCCATAGCCTTCCACTGCTGCC 1551
Qy 1081 CACGAGCTGGGCGAGGTTCACCATGCCCCATGAAAGTGTGAAAGTCTGTGAGGAGGTG 1140
Db 1552 CACGAGCTGGGCGAGGTTCACCATGCCCCATGAAAGTGTGAAAGTCTGTGAGGAGGTG 1611
Qy 1141 TTTGGGAAGCTCGGAGCAACACATGATGTCGCCGACCTCATCCAGATCGACCTGCTCC 1200
Db 1612 TTTGGGAAGCTCGGAGCAACACATGATGTCGCCGACCTCATCCAGATCGACCTGCTCC 1671
Qy 1201 AACCCCTGTGACGCTGCGAGTCTGCGATCATACCGACTTCTCGACAGCGGCGACGCT 1260
Db 1672 AACCCCTGTGACGCTGCGAGTCTGCGATCATACCGACTTCTCGACAGCGGCGACGCT 1731
Qy 1261 GACTGCTCTCTGGACCAACCCAGCAAGCCCATCTCCCTGCCCGAGGATCTTCCCGGGGCGC 1320
Db 1732 GACTGCTCTCTGGACCAACCCAGCAAGCCCATCTCCCTGCCCGAGGATCTTCCCGGGGCGC 1791
Qy 1321 AGCTACACCTGAGCGAGCTGCGAGCTGGGCTTTTGGGCTGGGCTTCCAGGCTTGTCT 1380
Db 1792 AGCTACACCTGAGCGAGCTGCGAGCTGGGCTTTTGGGCTGGGCTTCCAGGCTTGTCT 1851
Qy 1381 TACATCAGTACTGCAACCAAGCTGTGTGTCACCGGGAAGGCCAAGGACAGATGTTGTGC 1440
Db 1852 TACATCAGTACTGCAACCAAGCTGTGTGTCACCGGGAAGGCCAAGGACAGATGTTGTGC 1911
Qy 1441 CAGACCCGCGACTTCCCTTGGGCGGATGCGACAGCTGTGGCGAGGCGAAGCTCTGCTC 1500
Db 1912 CAGACCCGCGACTTCCCTTGGGCGGATGCGACAGCTGTGGCGAGGCGAAGCTCTGCTC 1971
Qy 1501 AAAGGGGCTTGGTGAGAGACACACCTTCAACAGCACAG----- 1541
Db 1972 AAAGGGGCTTGGTGAGAGACACACCTTCAACAGCACAG----- 2031
Qy 1542 -----GGTGGATGCT 1551
Db 2032 TCTCCAAAACACTCTTATTAAGGCTACCAATAGGCTGCACTACACAGGTGGATGCT 2091
Qy 1552 TCTTGGGCCAAATGGGATCCCTATGGCCCTTGTGCGGCACATGTTGGTGGGCGGTGCGAG 1611
Db 2092 TCCTGGGCCAAATGGGATCCCTATGGCCCTTGTGCGGCACATGTTGGTGGGCGGTGCGAG 2151

Qy	1612	CTGCCAGGAGGAGTGCACCAACCCACCCCTGCCAACGGGGGCAAGTACTCTCGAGGGA	1671
Db	2152	CTGCCAGGAGGAGTGCACCAACCCACCCCTGCCAACGGGGGCAAGTACTCTCGAGGGA	2211
Qy	1672	GTGAGGGTGAATACCGATCTTGCAACCTTGGAGCCCTGCCCCAGCTCAGCCTCCGGAAAG	1731
Db	2212	GTGAGGGTGAATACCGATCTTGCAACCTTGGAGCCCTGCCCCAGCTCAGCCTCCGGAAAG	2271
Qy	1732	AGCTTCCGGGAGGAGCAGTGCAGCGTTTCACGCGTTCACACACAGCACCAACCGGCTC	1791
Db	2272	AGCTTCCGGGAGGAGCAGTGCAGCGTTTCACGCGTTCACACACAGCACCAACCGGCTC	2331
Qy	1792	ACTCTCCCGTGCGATGGGTGCCCAAGTACTCCGCGGTCTCCCGGGAGCAAGTGCAG	1851
Db	2332	ACTCTCCCGTGCGATGGGTGCCCAAGTACTCCGCGGTCTCCCGGGAGCAAGTGCAG	2391
Qy	1852	CTCATCTCGGAGCCAAATGGCACTGGCTACTTCTATGTGTGGACCCCAAGGTGGTGAC	1911
Db	2392	CTCATCTCGGAGCCAAATGGCACTGGCTACTTCTATGTGTGGACCCCAAGGTGGTGAC	2451
Qy	1912	GGCACGCTGTGCTCTCTGACTCCACCTCCGCTCTGTGTCCAAAGCAAGTGCATCAAGGCT	1971
Db	2452	GGCACGCTGTGCTCTCTGACTCCACCTCCGCTCTGTGTCCAAAGCAAGTGCATCAAGGCT	2511
Qy	1972	GGCTGTGATGGAACTTGGGCTCCAAAGAGATTCGACAAAGTGTGGGGTGTGTGGGGGA	2031
Db	2512	GGCTGTGATGGAACTTGGGCTCCAAAGAGATTCGACAAAGTGTGGGGTGTGTGGGGGA	2071
Qy	2032	GACAAATAGAGCTGCAAGAAGTGACTGGACTCTTCACCAAGCCCATGCATGCAAT	2091
Db	2072	GACAAATAGAGCTGCAAGAAGTGACTGGACTCTTCACCAAGCCCATGCATGCAAT	2131
Qy	2092	TTGCTGGTGGCCATCCCGCAGGCGCTTCAAGCATCGACATCGCCAGCCGGTTACAAA	2151
Db	2132	TTGCTGGTGGCCATCCCGCAGGCGCTTCAAGCATCGACATCGCCAGCCGGTTACAAA	2191
Qy	2152	GGGCTGATCGGGATGACAATCTACCTGTCTGAAGAACCCCAAGCAAGTACCTGCTC	2211
Db	2192	GGGCTGATCGGGATGACAATCTACCTGTCTGAAGAACCCCAAGCAAGTACCTGCTC	2251
Qy	2212	AACGGGCATTTCTGGTGTCTGGCGGTGGAGCGGCACTGGTGTGAAGGCGCATCTGCTG	2271
Db	2252	AACGGGCATTTCTGGTGTCTGGCGGTGGAGCGGCACTGGTGTGAAGGCGCATCTGCTG	2291
Qy	2272	CGGTCACGGGCACGGGCACAGCGTGGAGCCTGCAAGCTTCCCGGGCCATCTCTGGAG	2331
Db	2312	CGGTCACGGGCACGGGCACAGCGTGGAGCCTGCAAGCTTCCCGGGCCATCTCTGGAG	2371
Qy	2332	CCGCTGACCGTGGAGGTCTCTCCGTGGGGAGATGACACCCGCCCGGGTCCGC	2385
Db	2372	CCGCTGACCGTGGAGGTCTCTCCGTGGGGAGATGACACCCGCCCGGGTCCGC	2395

```

RESULT 5
US-09-965-631-5
; Sequence 5, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Fridele, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: homo sapiens

```


Db 151 AGGTTTACAGCATCCAGCAATTAGGAATTCATTTAGCCTGGTGTGAGATCTTG 210
QY 811 CTCTTAGAGATGCTGACTCCGGGCCAAGGTCAACGGCAATGCGGCCCTGACGCTGCC 870
Db 211 GTCATATACGAGAGAGAGAGAGAGAGAGTTACCTCAATGCAGCTCTCACCTTCGG 270
QY 871 AACTTCTGTGCTGGCAGAGAGAGCTGAACAAAGTCAGTGACAGACACCCCGAGTACGG 930
Db 271 AATTTCTGACGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 330
QY 931 GACATGCCATCTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 990
Db 331 GACATGCCATCTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 390
QY 991 GGCATGGCTGATGCTGGTACCATGCTGACCCCAAGAGAGAGAGAGAGAGAGAGAG 1050
Db 391 GGAATGGCAGATGTTGGACCGTATGCTGACCCCAAGAGAGAGAGAGAGAGAGAG 450
QY 1051 GATGGCTTCCATFAGCCTTACACCTGCGCCACAGAGCTGGGCCACGCTGTTCACATGCC 1110
Db 451 GATGGCTTCCATFAGCCTTACACCTGCGCCACAGAGCTGGGCCACGCTGTTCACATGCC 510
QY 1111 CATGACAATGTGAAGCTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1170
Db 511 CAGATGATGCTGAAGCAGCTGTGCCAGCTTGAATGTGTGAGTGGGAGTTCTCATCTGATG 570
QY 1171 TCCCGGACCTCATTCAGATCGACCTGCAACCTGCTGACAGCTGCGAGTCTGCCATC 1230
Db 571 GCCTCGATGCTTCCAGCTTAGACCATAGCCAGCCCTGCTACCTTGCAGTGCCTACATG 630
QY 1231 ATCACCAGCTTCTTGACAGCGGGACAGTGTGCTCTGAGCAGCAGCCAGAGAGAGAG 1290
Db 631 GTCACGCTCCTCTCTAGATATGACAGCGGGAGATGTTGATGACAGCCCGAGAGTCCA 690
QY 1291 ATCTCCCTGCCAGAGATCGCGGGCCAGCTACACCTGAGCAGGAGTGCAGAGTG 1350
Db 691 ATCAAGCTCCCTTCTGATCTCCCGGTACCTTGTACGATGCCAAGCCGAGTGTGAGTTT 750
QY 1351 GCTTTTGGGTGGGTCTCAAGCCCTGCTCTTACATGCAGTAC---TGCACCAAGCTGTGG 1407
Db 751 ACATTCGGAGAGAGATCCAGCAGCTGCCCTGATGACAGCAGCAGATGACTACCTGTGG 810
QY 1408 TGCACCGGAG 1467
Db 811 TGCATGCGACCTCCGGTGGCTTACTGTGTGTGCTGCAAGTGGCAGATGCTTGGGAGAT 870
QY 1468 GGCACGAGCTGGGAGGAGAGAGCTTCCCTCAAGAGGAGCTGCGTGAGAGAGAGAGAG 1527
Db 871 GGCACGAGCTGGAGAGGAGAGAGTGTGTGCTGAGAGTGGAGTGGAGAGAGAGAGAG 930
QY 1528 CTCACAAGCAG 1581
Db 931 ATGAAGCATTTTGTCTACTCTCTTCAAGAGCTGGGAGCAGTGGGAGCAGTGGGAGAG 990
QY 1582 TGCCTGCGACATGCTGGGAG 1641
Db 991 TGCCTAAG 1050
QY 1642 CCGTCAAG 1701
Db 1051 CCAAG 1110
QY 1702 GAGCCCTGCGCAGCTAGCTCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1761
Db 1111 GAGGAGCTGCTCCAG 1167
QY 1762 AAGCGCTACACAG 1821
Db 1168 ATAGAGTTTCCAAAGCTTCTTTGGGAGATGAGCCCACTGTAGAGTGGACAGAGAGTAC 1227
QY 1822 TCCGGCGTGTCTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1881

Db 1228 GCCGGCGTCTCCCAAG 1287
QY 1882 TTCTATGTGTGGCAGCCCAAGGTGGTGAGCGCAGCGTGTCTCTCTGACTCCACCTCC 1941
Db 1288 TTTTTCGCTTTACAGCCCAAGGTGGTGAGCGCAGCTCCCTGTAGTCAGAGACTCTAC 1347
QY 1942 GTCTGTGTCCAAGGCAAGTGCATCAAGGCTGGCTGTGTGTGTGTGTGTGTGTGTGT 2001
Db 1348 GTCTGTGTGCAAGGCGAGTGTGTGAAAGCTGGCTGTGTGTGTGTGTGTGTGTGTGT 1407
QY 2002 AGATTTCGCAAGTGTGGGTGTGTGGGGAGAGACAATTAAGAGCTGCAAGAAGTCACTGGA 2061
Db 1408 AAGTTTGAATAGTGTGGGTGTGTGGAGAAACGGTTCACATGCAAGAAGTGTGAGGA 1467
QY 2062 CTCCTTCAAGAGCCCATGCATGGCTTACAAATTCGTGTGTGTGTGTGTGTGTGTGTGT 2121
Db 1468 ATAGTCACTAGTACAAGACCTGGGTATCATGACATTTGCAAAATTCCTGTGTGGAGCCACC 1527
QY 2122 AGCATCGACATCCCGCAGCGGTTCACAAAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2181
Db 1528 AACATTGAAGTGAACATCGGAATCAAAAGGGGTCCAGAAACAAATGGCAGCTTTCTGGCT 1587
QY 2182 CTGAAGAACAGCCAGGCAAGTACCTGTCTAACGGGCAATTCGTGTGTGTGTGTGTGTGTGTGT 2241
Db 1588 ATTAGAGCGCTGATGTGTACCTATATCTGAATGAAACTTCACCTGTCTCCACACTAGAG 1647
QY 2242 CGGACCTGTGTGTGAAGGGCAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2301
Db 1648 CAAGACCTCACCTACAAGGTACTGTCTTAAGGTACAGTGTGTGTGTGTGTGTGTGTGTGT 1707
QY 2302 AGCCTCGAGGCTTCCCGGCCATCTGTGAGCGCTGTGAGGCTGTGAGGCTGTCTCTCGCTGGG 2361
Db 1708 AGAATCCGAGCTTTAGTCCACTCAAGAGACCTTAACATCCAGGTCTTTATGGTAGGC 1767
QY 2362 AAGATGACAGCGCCCGGGTCCGCTACTCTCTATCTATCTGCCCCAAGAGCCTCGGGAGGAC 2421
Db 1768 CATGCTCTCCAGCCCAAAATTAATTCACCTACTTTATGAAGAAGAGACAGAGTCAATC 1827
QY 2422 AAGTCTCTC 2431
Db 1828 AACCCATTC 1837

RESULT 8

US-10-097-580-13

; Sequence 13, Application US/10097580

; Publication NO. US20030032168A1

; GENERAL INFORMATION:

; APPLICANT: Hirose, Kunitaka

; APPLICANT: Inoguchi, Eiji

; APPLICANT: Hakoziaki, Michinori

; APPLICANT: Ishioka, Keiko

; APPLICANT: Ishida, Yukako

; APPLICANT: Matsushima, Kouji

; APPLICANT: Kuno, Kouji

; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical

; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA

; FILE REFERENCE: 057092

; CURRENT APPLICATION NUMBER: US/10/097,580

; PRIOR FILING DATE: 2002-03-15

; PRIOR FILING DATE: 1999-12-03

; PRIOR APPLICATION NUMBER: JP 9-160422

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 13

; LENGTH: 2184

; TYPE: DNA

; ORGANISM: Mus sp.

; FEATURE:

; NAME/KEY: exon

; LOCATION: (1)..(2184)

US-10-097-580-13

Query Match 22.3%; Score 636; DB 9; Length 2184;
Best Local Similarity 50.7%; Pred. No. 2.8e-160;
Matches 1098; Conservative 0; Mismatches 700; Indels 12; Gaps 3;

QY 631 AAGGGTTTCGTCTATCCCGCGCTAGCTGGAGACGCTGGTGTCTCGCGACGAGTCAATG 690
DB 31 AAGGATTTGTCTCAGCCCGCTTATGTGAAACCATGCTCGTAGCTGACCACTCATG 90

QY 691 GTCAGTTTCCAGCGCGGAGCTGGAAATATCTGTGAGCGTGTGCGCAACGCGGCG 750
DB 91 GCCAGCTTCCAGCGGAGCGGTCTAAAGCATTAACCTCTTAACCTGTCTCGGTGGCAGC 150

QY 751 CGACTCTACCGCATCCAGCATCTCAACCCCATCAACATCTGTGTGTCAGAGTGTG 810
DB 151 AGGTTTACAGCATCCAGCATTAGGAATTAATAGCTGTGTGTTGTTGAAGATCTTG 210

QY 811 CTTCTTAGACTCTGTACTCCGGGCCCAAGGTCAACCGCAATCGGCCCTGACGCTGCG 870
DB 211 GTCATATACGAGGAGAGAGGACCAAGATTACCTTCAATCGACTCTCAACCTTCGG 270

QY 871 AACTTCTGTGCTGGAGAGAGCTGAACAAAGTGTGAGTGTGAGCAAGCAACCCGAGTACTG 930
DB 271 AATTCTGCACTGGCAGTGGCAGAACACACACAGCCCGCAGTACCGGGATCCAGAGCACTAT 330

QY 931 GACACTGCACTCTTCCAGCAGGAGCTGTGTGGAGCCACCACTGTGTGACACCCCTG 990
DB 331 GACACTGCAATCTGTTCACAGACAGATTTATGTGGCTCCACAGGTGTGACACTCTC 390

QY 991 GGCATGCTGTATGGGTACCATGTGTGACCCCAAGAGAGTGTCTGTCTGATGTAGGAC 1050
DB 391 GGAATGGCAGATGTGGAACCGTATGTGACCCAGCAGGAGCTGCTCAGTCAATAGAAT 450

QY 1051 GATGGCTTCCATCAGCTTCCACACTGCCAGAGCTGGCCACGCTGTTTCAACATGCC 1110
DB 451 GATGGTTGCAAGTGTCTTCCACACAGCCCATGAATGTGGCCATGTGTTTACATGCC 510

QY 1111 CATGACATGTGAAGTCTGTGAGGAGGTGTTGGGAAGTCTCGAGCCCAACCATGATG 1170
DB 511 CACGATGATGTAAGCACTGTGCCAGTGTGAATGGTGTGAGTGGCGATTTCTCATCTGATG 570

QY 1171 TCCCGGACCTCATCAGATGCGCCGTGCCAACCCCTGGTTCAGCTCGAGTGTGCCATC 1230
DB 571 GCCTCGATGTCTCCAGCTTAGACCATAGCCAGCCCTGGTCACTTCGAGTGCCTACATG 630

QY 1231 ATCACCAGTTCCTGGAGCGGCGAGTGTGACTGCTCTCGGACCAACCCAGCAAGCCC 1290
DB 631 GTCAGTCTTCCCTAGATATATGACACGCGGAATGTTGTGAGCAAGCCCCAGATCCA 690

QY 1291 ATCTCCCTGCGGAGGATCTGCGCGGCGCCAGTGTACACCTGTAGCCAGCAGTGGAGCTG 1350
DB 691 ATCAAGCTCCCTTCTGATCTTCCCGGTACCTTGTACGATGCCAACCCGCAAGTGTGAGTTT 750

QY 1351 GCTTTTGGCGTGGCTCCAACCCCTGTCTTACATGACGTAC ---TGCACCAAGCTGTGG 1407
DB 751 ACATTCGGAGAGGAATCCAGCATGCGCCCTGATGACGACGACATGTACTACCTCTGG 810

QY 1408 TGCACCGGAAGGCGAAGGAGACAGATGGTGTGCCAGACCCCGCCACTTCCCTCGGCGCAT 1467
DB 811 TGCAGTGGCACTCGGCTGGCTTACTGGTGTGCCAACAACAACTTCCCTTGGCGAGAT 870

QY 1468 GGCACAGCTGTGGCGGAGGAAAGCTGTGCTCAAGAGGGGCTCGGTGGAGAGACACAAC 1527
DB 871 GGCACAGCTGTGGAGAGGGAAGTGTGTGTCAGTGGCAAGTGGTGAACAGACAGAC 930

QY 1528 CTCACACAGCAGACAG ---GCTGGATGTTCTTGGCCCAATGGATCCCTATGGCCCC 1591
DB 931 ATGAAGCATTTTGTCTTACTCTCTGTTATGAGAGCTGGGACCATGGGACCGTGGGAGAC 990

QY 1592 TGCTCGGCACATGTGGTGGGGGGGTGACGTGGCCAGGAGGAGTGCACCAACCCCAACC 1641
DB 991 TGCTCAAGAACCTGTGGTGGTGGAGTTCAATACAAATGAGAGAAATGTGCAACCCAGTC 1050

QY 1642 CCTGCAACGGGGCAAGTACTCTGGAGGAGTGGAGGTGAATAATCCGATCTCTGCAACCTG 1701
DB 1051 CCAAGAAACGGAGGCAAGTACTGTGAAGGCAAAACAGTCCGCTACAGGTCTCTGTAACATC 1110

QY 1702 GAGCCTCGCCCGCAGCTCAGCCTCCGGAAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTC 1761
DB 1111 GAGGACTCTCCAGACAAATAA ---CGGAAACAGTTTCAGAGAGGAGCAGTGTGAGGCGCAC 1167

QY 1762 AAGGCTACAACACACAGCACCAACCGGCTCACTCTCGCCGTGGATGGTGGCTGCCCAAGTAC 1821
DB 1168 AATGAGTTTCCAAAGCTTCTTTGGGAATGAGCCCACTGTAGAGTGGACACCCCAAGTAC 1227

QY 1822 TCCGCGTGTCTCCCGGCAAGTGCAGAGTCACTCTCCGAGCCAAATGGCACTGGCTTAC 1881
DB 1228 GCGCGCTCTCCCAAGGACAGGTGCAAGCTCACTCTGGAAGCCAAAGGCAATTTGGCTTAC 1287

QY 1882 TTCTATGTGCTGGCAGCCCAAGGTGGTGGAGCGGACGCTGTCTCTCTGACTCCACCTCC 1941
DB 1288 TTTTTCGCTTTACAGCCCAAGGTTGTAGATGCACTCCCTGTAGTCCAGACTCTTACCTCT 1347

QY 1942 GTCTGTGTCCAAGCAAGTGCATCAAGGCTGGCTGTGATGGGAACCTGGGCTCCCAAGAA 2001
DB 1348 GTCTGTGTCAAGGCGAGTGTGTGAAGCTGGCTGTGATCGCATCATAGACTTCCAAAAAG 1407

QY 2002 AGATTGCAAGAGTGTGGGTGTGTGGGGAGACAAATAAGAGCTGCAAGAGTGTACTGGA 2061
DB 1408 AAGTTTGTAGTGTGGGCTTTGTGGAGGAAACGGTTCCACATGCAAGAAGATGTGACGA 1467

QY 2062 CTTCTCACCAGCCCATGCTGCTACAAATTTCTGTGGTGGCCATCCCGCAGGCGCTCA 2121
DB 1468 ATAGTCACTAGTACAGAGCTGGGTATCATGACATGTGCAATTTCTTCTGCTGGAGCCACC 1527

QY 2122 AGCATGCACTCGCCAGCGGTTTACAAAGGCTGTATCGGGATGACAACTTACCTGGCT 2181
DB 1528 AACATTGAGTGAACATCGGAATCAAAGGGGTCCAGAAACATGCGAGCTTCTGGCT 1587

QY 2182 CTGAGAACAGCCAGGCAAGTACCTGCTCAACGGGCAATTCGTGTGCTGGCGGTGGAG 2241
DB 1588 APTAGAGCGCTGATGGTACCTATATTCTGAATGGAACCTTCACTCTGCTCCACACTAGAG 1647

QY 2242 CGGAGCTGTGTGAAGGCGAGTCTGTGCGGTACAGCGGCAGCGGACAGCGTGGAG 2301
DB 1648 CAAGACTCACCTACAAAGTACTGTCTTAAGTACAGTGGTTCTCGGCTCGCTGGAA 1707

QY 2302 AGCCTGCAAGCTTCCCGGCCCATCTGTGGAGCGCTGACCGTGGAGGTCTCTCCGTGGGG 2361
DB 1708 AGAATCCGCAAGCTTTAGTCCACTCAAAAGAACCCCTTAACCATCCAGGTTCTTATGTTAGGC 1767

QY 2362 AAGATGACACCGCCCGGCTCGCTACTCTTCTATCTGCCCAAGAGCCTCGGGAGGAC 2421
DB 1768 CATGCTCTCGACCCCAAAATTAATTTACCTACTTTATGAAGAAGACAGAGTCATTC 1827

QY 2422 AAGTCCCTCTC 2431
DB 1828 AACGCCATTC 1837

RESULT 9
US-09-445-023A-13
; Sequence 13, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakoizaki, Michinori
; APPLICANT: Ishioke, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092

Sequence 2, Application US/10097597
Publication No. US20030022352A1

GENERAL INFORMATION:

APPLICANT: Hirose, Kunitaka
APPLICANT: Inoguchi, Eiji
APPLICANT: Hakozaaki, Michinori
APPLICANT: Ishioaka, Keiko
APPLICANT: Ishida, Yukako
APPLICANT: Matsushima, Kouji
APPLICANT: Kuno, Kouji

TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,

TITLE OF INVENTION: pharmaceutical

TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS

FILE REFERENCE: Q57092

CURRENT APPLICATION NUMBER: US/10/097,597

CURRENT FILING DATE: 2002-03-15

PRIOR APPLICATION NUMBER: 09/445,023

PRIOR FILING DATE: 1999-12-03

PRIOR APPLICATION NUMBER: JP 9-160422

PRIOR FILING DATE: 1997-06-03

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn version 3.0

SEQ ID NO 2

LENGTH: 2184

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: exon

LOCATION: (1)..(2184)

US-10-097-597-2

Query Match 22.0%; Score 628.8; DB 9; Length 2184;

Best Local Similarity 60.9%; Pred. No. 2.4e-158;

Matches 1080; Conservative 0; Mismatches 682; Indels 12; Gaps 3;

QY 631 AAGGTTTCGTCTATCCCGCGGTAGCTGGAGACGCTGTGTGTCGGGACGAGTCAATG 690
DB 31 AAGCGATTGTGTCCAGCCCGGTATGTGGAAACCATGCTTGTGGCAGACCACTCGATG 90
QY 691 GTCAAGTTCACAGCGCGGACCTGGACATATTCTGTGACGCTGTGTGGCAACGCGGCG 750
DB 91 GCAGAAATCCAGCGCAGTGTGCTAAAGCAATACCTCTCAGCTGTGTTTCGGTGGCAGC 150
QY 751 CGACTACCGCCATCCAGCATCTCAACCCATCAACATCTGTTGGTCAAGTGTCTG 810
DB 151 AGATTGTACAAACCCAGCAGCTCGTAATTCAGTTAGCTGTGTGTGGTGAAGATCTTTG 210
QY 811 CTTCCTTAGAGATCGTGACTCCGGGCCCAAGGTACCGGCAATCGGCCCTGACGCTGCGC 870
DB 211 GTCATCCAGATGAACAGAGGGCGCGGAAGTGAACCTCAATGCTGCCCTCACTCTGGCG 270
QY 871 AACTTCTGTGCTGGCAGAAAGAGCTGGAACAAAGTGAAGTGAACAGCAACCCGAGTACTGG 930
DB 271 AACTTTTGCACCTGGCAGAGAGCAGCAACCCAGTCAACCGGATGCGAGGCACTAT 330
QY 931 GACACTGCCATCTTCCAGCAGGAGCTGTGTGGAGCCAGCAGCTGTGACACCCCTG 990
DB 331 GACACGAATATCTTTTCCACGACAGAGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 390
QY 991 GGCATGCTGTGTGGTGTGATGTGTGACCCCAAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 1050
DB 391 GGGATGCTGT 450
QY 1051 GATGGGCTTCCATCAGCTTTCACCATGTGCCCCAGAGCTGGGCCAGCTGTTTCAACATGCC 1110
DB 451 GATGGTTTACAAAGCTGCTTCCACAGCCCATGAATAGGCGCCAGCTGTTTAAACATGCCA 510
QY 1111 CATGACATGTGAAGTCTGTGAGGAGTGTGTTGGGAGCTTCCGAGCCACCATGATG 1170
DB 511 CATGATGATGCAAGAGCTGTGCGAGCTTAAATGGTGTGAACAGGATTTCCACATGATG 570
QY 1171 TCCCGGACCTCATCCAGATCGACCGTGGCCAAACCCCTGGTGCAGCTGCAGCTGCATC 1230

DB 571 GGGTCAATGCTTTCCAAACCTGGACACAGCCAGCGCTTGGTCTCCTTGGAGTGCCTACATG 630
QY 1231 ATCACCGACTTCTTGGACAGCGGCGACGGTGAAGTCTCTTGGACCAACCCAGCAAGCC 1290
DB 631 ATTACATCATTTCTGGATAATGGTCAATGGGAATGTTTGGATGGACAAGCCTTCAGATCCC 690
QY 1291 ATCTCCCTGCGCGAGGATCTCCGCGGCGCCAGCTACACCTGTAGCCAGCAGCAGCTG 1350
DB 691 ATACAGCTCCCGAGCGATCTCCCTGGCACCTTGTACAGTGCACACCGGCGAGTGCAGTTT 750
QY 1351 GCTTTTGGCTGGCTCCCAAGCCCTGCTTACATGCAATGACAGTAC---TGCAACCAAGCTTGG 1407
DB 751 ACATTTGGGAGGACTCCAAACACTGCCCCGATGAGCCAGCAGATGTAGCAGCTTTGG 810
QY 1408 TGCAACCGGAAGGCAAGGAGCAGATGTGTCCAGAGCCCGCCACTTCCCTGGGCGCAT 1467
DB 811 TGTACCGGACCTCTGGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 870
QY 1468 GGCACACAGCTGTGGGAGGGAAGCTCTGCTCAAGGGGCGCTGCTGGAGAGACACAAAC 1527
DB 871 GGCACACAGCTGTGGAGAGGGAATGTTATCAACGGCAAGTGTGTGAACAAACACGAC 930
QY 1528 CTCACAAAGCACAGG-----GTGGATGTTCTTGGGCCAAATGGGATCCCTATGGGCC 1581
DB 931 AGGAAGCATTTTGTATACGCTTTTTCATGGAAGCTGGGAGCCATGGGGACCGTGGGAGAC 990
QY 1582 TGCTCGGCACATGTGTGGGCGGTGCAGCTGGCCAGGAGGAGTGCACCAACCCAGC 1641
DB 991 TGTTCGAGAACGCTGGGCTGGAGAGTCCAGTACACGATGAGGAATGTGACAAACCCAGTC 1050
QY 1642 CCGTCCACAGGGGCAAGTACTGCGAGGAGTGAAGGTGAATACCGATCTGTAACCTG 1701
DB 1051 CCAAGAAATGGAGGAAGTACTGTGAAGGCAACAGAGTGCAGATCTGTAACTT 1110
QY 1702 GAGCCCTGCCCCAGCTCAGCCCTCCGGAAGAGCTTCCGGGAGGAGCAGTGTGAGCTTTC 1761
DB 1111 GAGGACTGTCCAGC---AATATGGAAGAACCTTTAGAGAGGAACATGTGAAGCACAC 1167
QY 1762 AACGGCTACAACCAAGCAGCAGCCGCTACTCTGCGCGGTGGATGGTGGCCCAAGTAC 1821
DB 1168 AACGAGTTTTCAAAAGCTTCTTTTGGGAGTGGGCGCTGGGAGTGAATGGAATGCCAAGTAC 1227
QY 1822 TCCGCGTGTCTCCCGGAGCAAGTGAAGTCAATCTCCGAGCAATGCGACTGGCTAC 1881
DB 1228 GCTGGCTCTCAACAAAGACAGGTGCAAGTCACTCTGCCAAGCCAAAGGATTTGGCTAC 1287
QY 1882 TTCTATGTCTGGCACCCCAAGTGTGTGACGCGACGCTGTGCTCTCTGATCTCCACTTCC 1941
DB 1288 TTCTTCGTTTTCAGCCCAAGGTTGTTGATGTTACTTCCATGTAGCCAGATTTCCACTCT 1347
QY 1942 GTCTGTGTCAGGCAAGTGCATCAAGCTGGCTGTGTGAGGAACTGGGCTGCCAAGAG 2001
DB 1348 GTCTGTGTGCAAGGACAGTGTGTAAAGCTGGTGTGTGATCGCATATAGATCTCCAAAAG 1407
QY 2002 AGATTGCAAGTGTGGGTGTGTGGGAGACAAATAAGAGCTGCAAGAAAGTGAAGTGA 2061
DB 1408 AAGTTGATAATGTGGTGTGGGGAATGATCTACTTGTAAATAATATACAGTA 1467
QY 2062 CTCTTCAACCAAGCCATGATGGCTACAATTTCTGTGGTGGCCATCCCGCAGGCGCTCA 2121
DB 1468 TCAGTTACTAGTGCAAAACCTGGATATCATGATATCGTCACAATTTCCAACTGGAGCCACC 1527
QY 2122 AGCATGACATCCCGCAGCGGTTTACAAGGCTGTAGGGGATGCAACATACCTGCT 2181
DB 1528 AACATCGAAGTGAACAGCGGAACAGAGGGGATCCAGGAACAATGGCAGCTTTCTGCC 1587
QY 2182 CTGAAGAACAGCAAGGCAAGTACCTGCTCAACGGGCAATTTCTGTGTGCGCGGTGGAG 2241
DB 1588 ATCAAGCTGCTGATGGCACATATATCTTAATGTGTACTACACTTTTCTCCACTTAGAG 1647
QY 2242 CGGGACCTTGTGTGAAGGCGAGTCTGTGCGGTACAGCGGACGGGACAGCGGTGGAG 2301
DB 1648 CAAGACATTATGTACAAAGGTTGTGTCTGTAGGTACAGCGGCTCTCTGCGGCAATGGAA 1707

Qy	2302	AGCCTGCAGCGTTC	CCCGCGGCCATCT	CTGGAGCGCTG	ACCGTGGAGGTC	CTCTCCGTGGG	2361
Db	1708	AGAAATCGCAGCTT	ATGACCTCTCA	AAGAGCCCTT	GACCATCCAGG	TCTTACTCTGGC	1767
Qy	2362	AAATGACACCGCC	CGCCGGGTCG	CGCTACTCTCT			
Db	1768	AATGCCCTTCG	ACCTAAAAT	TAAATACACCT	ACT		

RESULT 11

```

RESULT 11
US-10-097-580-2
; Sequence 2, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiiji
; APPLICANT: Hakozaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; FILE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2184
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)..(2184)
US-10-097-580-2

```

Query Match	22.0%;	Score 628.8;	DB 9;	Length 2184;
Best Local Similarity	60.9%;	Pred. NO. 2.4e-158;		
Matches 1080;	Conservative	0;	Mismatches 682;	Indels 12;
				Gaps

QY 631 AAGCGTTTCGTGCTATCCCGGGTACGTGCAGACGCTGGTGTCGGGACGAGTCAATG 690
||||| ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
Dd 31 AAGCGATTGTGTCCAGCCCCCCTTATGTGGAACCATGCTCTGGCACCACTCGATG 90

[illegible][illegible]

QY 811 CTCTTTAGAGATCGTGACTCCGGGCCCCAAGTCACTGGGCAATCGGGCCCTGACGCTGGCG 870

871 AAC TTC TGT GCG CAG AAG AGT GAC AAA AGT GAG T GAC AAG CAC CCC GAG TACT GG 930

DD	2/1	AAC	TTTTT	GCA	CTGGC	AGA	AGC	GAC	CA	CC	CAC	CC	CAG	TG	AC	CGG	GAT	GC	AG	GC	ACT	AT	330
QY	931	GAC	ACT	GCC	ATC	CTT	CAC	CAG	GAC	CTG	TG	TG	GAG	CC	ACC	CA	CTT	GT	G	AC	ACC	TTG	990

Db	331	GACACAGCAATCTTTTACCACAGACAGGACTGTGTGGGTCACAGACATGTGATACTCTT	390
QY	991	GGCATGGCTGATGTGGGTACCATTGTGTGACCCACAGAGAAGTGCTCTGTCTATTGAGGAC	1055

Db 391 GGGATGGCTGATGTTGGAACTGTCTGTGTATCCGAGCAGAAGCTGCTCGTTCATAGAAGAT 450

Qy	1051	GATGGCTTCATCAGCCCTTCA	CCACTTGC	CCACAGCTGGCCACAGCTGGCCACATGCC	1111
Db	451	GATGGTTACAGCTGCTTCA	CCACAGCCCAATGAATAGGCACAGTGT	TTAACTGCA	510
Qy	1111	CATGCAATGAAAGTCTCTG	GAGGAGTGT	TTGGGAAGCTCCGAGCCAAACCATGATG	1170
Db	511	CATGATGATGCAAGCAGTGT	GCCAGCTTAA	TGTTGGTGTGAACACAGGATTC	570
Qy	1171	TCCCGACCCCTCATCCAGAT	TCGACCGTGGC	CAACCCCTGGTCTAGCTGCGAGTCTGCCAT	1230
Db	571	CGGTCAATGCTTTCCAACT	TGGACACAGCCAGCCTTGGTCTCCTTGCAGTGCCTACATG	630	
Qy	1231	ATCACCGACTTCTCTGGAC	ACGGGCGAGTGTACTGCTCTCTGGACCAACCCAGCAAGCC	1290	
Db	631	ATTACATCATTTCTGGATA	ATGTCTATGGGAATGTTTGA	TGGACAAGCCCTCAGAAATCCC	690
Qy	1291	ATCTCCCTGCCGAGGATCT	CGCGGGCGCAGCTACACCCTGAGCAGCAGTGGCAGGTG	1350	
Db	691	ATACAGCTCCAGCGCATCT	CCCTGGCACCTTGTAGGATGCCAACCGCAGTGGCAGTTT	750	
Qy	1351	GCTTTTGGCTGGGCTCCAA	AGCCCTGTCTCTACATGCAAGTAC---TGACACAAAGCTTTGG	1407	
Db	751	ACATTTGGGAGGACTCCA	AACACTGCCCGGATGACCGCAGCAGCATGTAGTACCTTGTGG	810	
Qy	1408	TGCACGGGAAGGCCAAGG	CACAGATGTGTCCAGACCCGCCAGCTCCCTCGGCGCCGAT	1467	
Db	811	TGTACCGGCACTCTGGT	GGGTGCTGGTGTCTAAACCAACACTTCCCTGGGCGGAT	870	
Qy	1468	GGCACCAAGCTGTGGCAGG	GCACAGCTCTGCCTCAAGGGGCGCTGGTGGAGAGACACAAC	1527	
Db	871	GGCACCAAGCTGTGGAGA	AGGAAATGGTGTATCAACGGCAGTGTGTGAACAAACCCGAC	930	
Qy	1528	CTCAACAGCACAGG-----	GTGGATGTTCTCTGGGCGCAATGGGATCCCTATGGCCCC	1581	
Db	931	AGGAAGCATTTTGATAC	GCCTTTTCATGAAGCTGGGACCATGGGGACCTGGGGAGAC	990	
- Qy	1582	TGCTCGCGCACATGTGTG	GGGCGGTGCAGCTGGCCAGGAGCAGTGCACCAACCCCAACC	1641	
Db	991	TGTTTCGAGAAGCTGCG	GTGGAGAGTCCAGTACACGATGAGGAAATGTGACACCCAGTC	1050	
Qy	1642	CCTGCCAACCGGGGCAAG	TACTCGCAGGAGTGGAGGTGAATACCGATCTCGCAACCTG	1701	
Db	1051	CCAAAGAATGGAGGGAAG	TACTGTGAAGCAACAGAGTCGCTACAGATCTCTGTAACTT	1110	
Qy	1702	GAGCCCTTCCCCAGCTCAG	CCCTCCGGAAGACTTCCGGGAGGAGCAGTGTGAGGCTTTC	1761	
Db	1111	GAGGACTGTCCAGAC--	---AATATGGAACAACTTTAGAGGAGACATGTGAAGCACAC	1167	
Qy	1762	AACGGCTTCAACACAGC	ACCAACCGGCTCCTCTCGCGGTGGCATGGGTGCCCAAGTAC	1821	
Db	1168	AACAGGTTTTCAAAAGCT	TCCCTTGGGAGTGGGCGCTGGGTGGAAATGGATTTCCCAAGTAC	1227	
Qy	1822	TCCGGCGTGTCTCCCGG	GACAGTCAAGCTCATCTGCCGAGCCAAATGGCACTGGCTAC	1881	
Db	1228	GCTGGCTCTCACCAAG	CAGAGTGCATCTGCCAAGCCAAAGGCAATTTGGGTACTTCCATGTAGCCAGATTTCCACCTCT	1347	
Qy	1882	TTCTATGTCTGGCACCC	CAAGTGTGTGGACGCACGCTGTCTCTCTGACTCCACCTCC	1941	
Db	1288	TTCTTCTTTTGCAGCC	CAAGGTTGTGATGTTACTTCCATGTAGCCAGATTTCCACCTCT	1347	
Qy	1942	GTCTGTTCCAAGCAAGT	GCATCAAGCTGCTGTGATGGGAACCTTGGGCTCAAGAAG	2001	
Db	1348	GTCTGTGTCAAGGACAG	TGTGTAAAGCTGGTTGTGATCGCATCTATAGCTCCAAAG	1407	
Qy	2002	AGATTCCACAAGTGTGG	GTGTGGGGGAGACAATAAGAGCTGCAAGAAGGTGACTGGA	2061	
Db	1408	AAGTTTGATAAATGTG	TTGTTGGGGGAATGATCTACTTGTGAATAAATAATCAGGA	1467	
Qy	2062	CTCTTCAACAGCCCATG	CATGGCTACAATTTCTGTGGTGGCCATCCCGCAGGCGCTCA	2121	
Db	1468	TCAGTTACTAGTCAAA	ACCTGGATATCATGATATCGTCACAATTTCAACTTGGAGCGAC	1527	

QY 2002 AGATTGACAAAGTGTGGGGTGTGTGGGGAGACAATAAGAGCTGCAAGAGGTGACTGGA 2061
Db 1408 AAGTTTCATTAATGTGTGTGGGGGAAATGGATCTACTTGTAAAAAATATCAGGA 1467
QY 2062 CTCTTACCAAGCCATGCTATCAATTTCTGTGGTGGCCATCCCGCAGCGCCTCA 2121
Db 1468 TCAGTTACTAGTCAAAACCTGGATATCATGATATCGTCACAAATTCAACTGAGCCACC 1527
QY 2122 AGCATGACATCCCGCAGCGGTTTACAAGGCTGATCGGGATGACAACTACCTGGCT 2181
Db 1528 ACATCGAAGTGAACAGCGGAAACAGAGGGGATCCAGGAACAATGCGAGCTTCTTGGC 1587
QY 2182 CTGAAGAACAAGCAAGTACCTGCTCAACGGGCATTTGCTGTGTGGGGTGGAG 2241
Db 1588 ATCAAGCTCTGATGSCACATATTTCTTAATGTGTACTACACTTTGTCCACCTTAGAG 1647
QY 2242 CGGAGCTGTGTGAAGGCGAGTGTGCTGCGGTAGACGGGACAGGGGACAGGGTGGAG 2301
Db 1648 CAAGACATTAATGACAAAGGTGTGTGTTGAGGTACAGCGGCTCTCTCGGCAATTGGA 1707
QY 2302 AGCTGAGGCTTCCCGGCGCATCTCGGAGCGCTGACGCTGGAGTCTCTCGGTGGG 2361
Db 1708 AGAATCGAGCTTTAGCCCTCTCAAGAGCCCTTGACCATCCAGTCTTACTGTGGGC 1767
QY 2362 AAGATGACACCGCCCGGGTCCGCTACTCTCT 2395
Db 1768 AATGCCCTTCGAGCTAAATAAATACACCTACT 1801

RESULT 13
US-10-105-929-1
; Sequence 1, Application US/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US/10/105,929
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4676
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (460)...(3360)
US-10-105-929-1

Query Match 22.0%; Score 628.2; DB 12; Length 4676;
Best Local Similarity 57.1%; Pred. No. 3.9e-158;
Matches 1305; Conservative 0; Mismatches 913; Indels 69; Gaps 6;

QY 175 GCATTTCAGGAGGACTTTTACCTACACCTGACGCGCGGATGCTCACTTTGGTCCCGCC 234
Db 697 GCGTTTGACGACGCTGGATCTGGAGCTGGGCGCCACAGTAGCTTTTGGCGCCCGC 756
QY 235 TTCTCCACTGACATCTGGGCTGCCCTCCAGGGGCTCACCGGGGCTC--TTCAGAC 291
Db 757 TTCAGCTCCAGACGTGGGGCGCAATCCGGGTCCGAGAGCCCTTCCGGAACCCGAC 816
QY 292 CTGCGACGCTGTTTATCTGGGACGTGAACGCCGCGGAGTCTTCTGCTGCTGTG 351
Db 817 CTGGCGCACTGTCTTACTCCGGCACCGTGAATGCGCATCCAGCTCGGCTGCCCGCC 876

QY 352 AGCCTGTGGGGGGCTCCGCGGAGCCCTTTGGCTACCGAGGGCGCGAGTATGTCTATTAGC 411
Db 877 AGCCTCTCGGAGGGCGTGGCGGCGCCCTTCTACCTGCTGGGGAGGCGCTATTATTCACG 936
QY 412 CGCTGCCCCAATGCTAGCGCGC-CGGCGCGCAGCGCAACAGCCAGCGCGCACCATCTCT 470
Db 937 CGCTGCCCCCGCGCAGCAGGCGCTCCACCGCGCCCGCCAGGGGAGAAAGCCGCGGCA 996
QY 471 CCAGCGCGGGGTGTTCGGGCGGGCTTCGGGAGACCCACCTCTCTCGCTCGGGGGTGGC 530
Db 997 CCACCTACAGTTCCACCTCTCGCGCGGAATCCGCGAGGCGACGTAGCGCGCACCTGCGGG 1056
QY 531 CTCGGGCTTGAACCCCGCATCTTACGGGCGCTTGACGCCCTTACAGCGCGCGCGGGCGG 590
Db 1057 GTCTGTGGACGAGCGCCCGCGGCTGCGAAAGCGGAGACCGAGAGCGAGGAGGAAAGG 1116
QY 591 CTTCCGGGAGAGTCTGATCCGCGCGCAGGCTCTGGGCGC----- 627
Db 1117 ACTGAGGCGGAGGAGGAGGCGCTCAGTGTGTCGCGGAGGACCCGGCACCTCAAGGGCTA 1176
QY 628 -----GCCAAGCGTTCGTGTCTATCCCGCGGTAC 657
Db 1177 GCACAGCCACAGGAACCTGGAAGCATAGAAAGAGGATTTGTCTCCAGTACCGCTAT 1236
QY 658 GTGGAGACGCTGTGTGGCGGACGAGTCAATGGTTCAGTTCCAGGCGCGGACCTGGAA 717
Db 1237 GTGNAACCATGCTGTGGCAGACCATGTCATGTCGAGATTCACGCGCAGTGGTCTAAAG 1296
QY 718 CATTATCTGTGACGCTGTCTGGCAACGCGCGGCTCTACCGCCATCCCGACGATCCTC 777
Db 1297 CATTACCTTCTCACGTTGTTTTCGCTGCGACGACAGATTGTACAAACACCCAGCATTCG 1356
QY 778 AACCCCATCAACATCGTTGTGGTCAAGTGTCTTCTTGTAGATCTGTGCTCCGCGCC 837
Db 1357 AATTCACTGTAGCTGTGTGGTGAAGATCTTGTGTCATCCACGATGAACAGAGGGCGC 1416
QY 838 AAGTCTACCGCAATGCGGCGCTGACGCTGCGCAACTTCTGTGCTGCGAGAAAGCTG 897
Db 1417 GAAGTACCTCCATGTCGCTCTCTGCGGAACCTTTGCAACTGGCAGAGACAC 1476
QY 898 AACAAAGTGAAGACAGACCCCGAGTACTGGGACACTGCCATCTCTTCCACAGCAG 957
Db 1477 AACCCACCCAGTACCGGGATGACAGACACTATGACACAGCAATCTTTTCAACAGAC 1536
QY 958 GACCTGTGTGAGCACCACCTGTGACACCTGGCATGGCTGATGTGGTACCATGTGT 1017
Db 1537 GACTGTGTGGTCCAGACATGTGATGATCTTGGGATGGCTGATGTGGAACGTGTGT 1596
QY 1018 GACCCCAAGAGAAGCTCTGTCTATTGAGACGATGGGCTTCCATCAGGCTTCAACACT 1077
Db 1597 GATCCGAGCAGAAGCTGCTCGCTCATAGAAGATGATGTTTACAAGCTGCTTCCACACA 1656
QY 1078 GCGCACGAGCTGGGCGCAGCTGTTCAACATGCCCATGACAAATGTGAAGTCTGTGAGG 1137
Db 1657 GCCCATGAATTAGGCCACGTTTAAACATGCCACATGATGATGCAAGAGTGTGCCAGC 1716
QY 1138 GTGTTTGGGAAGCTCCGAGCACAACCATGATGTCGCCGACCTCATCCAGATGCGACCT 1197
Db 1717 CTTAATGCTGAACACGAGATTCCCATGATGGCGTCAATGCTTTTCCAACCTGGACAC 1776
QY 1198 GCCAACCCCTGGTCAAGCTGCTGCTGCTGCTCATCATCCGACTTCTTGGAGAGGGCAC 1257
Db 1777 AGCCAGCCTTGGTCTCCTTTCAGTGCCTACATGATTAACATCTTCTGTGATTAATGGT 1836
QY 1258 GGTGACTGCTCTTGGACCAACCCAGCAAGCCCATCTCCCTGCCCGAGGATCTGCCGGC 1317
Db 1837 GGGGAATGTTGATGGACAAGCCTTCAGATTCCTCATACAGCTCCCGAGCGCATCTCCCTGC 1896
QY 1318 GCCAGCTACACCTTGAGCCAGCAGTGGAGCTGGCTTTTGGCGTGGCTCCAAAGCCTGT 1377
Db 1897 ACCTCGTACGATGCCAACCGGCGAGTGCACGTTTACATTTGGGAGGAGACTCCAAACACTGC 1956

Db 1037 CTGAGAACACATCTCTCAGGTGATGTCATAGGACCCGATCTACAGCACCCGAC 1096
QY 772 ATCTCAACCCATCAACATCTGTTGGTCAAGGTGCTGTTCTTAGAGATCGTGACTCC 831
Db 1097 ATCAGGAACCTCCGTCAACCTTGTGGTGGTGAAGTCTAATAGTGGAAAGAAAGATGG 1156
QY 832 GGCCCAAGGTCAACCGCAATGGGCCCTGACGCTCGCAACTTCTGTGCTCGGAGAG 891
Db 1157 GGCCGGAAGTGTCCGACACACGGGGGCTCACACTGCGCACTTCGACGTGGCAACGG 1216
QY 892 AAGCTGAACAAAGTGTGACAAAGACCCCGAGTACTGGGACACTGCCATCTCTTCAAC 951
Db 1217 CGTTTCAACAAGCCCAAGTACCGCCACCGGAGCACTATGACACTGCCATCTTGTTCAC 1276
QY 952 AGCAGGACCTGTGTGG---AGCCACCACTGTGACACCCCTGGGATGGCTGATGGGT 1008
Db 1277 AGACAAACTTGTGGGAAGGAGAGAGTGTGACACCCCTGGGATGGCAGACCTTGGC 1336
QY 1009 ACCATGTGTACCCCAAGAAAGTCTCTGTCTATTGAGGAGGATGGGCTTCCATCAGCC 1068
Db 1337 ACCATGTGTACCCCGACAAAGAGTCTCTCAGTATCAAGATGAGGACTGAGCAGCC 1396
QY 1069 TTCACCACTGCCACAGCTGGCCACGTTTCAACATGCCCATGCCATGTGAAGTCC 1128
Db 1397 TACACCCCTGGCCCATAGCTAGGACGCTTCTCAGCATGCCCATGATTTCTAAGGCC 1456
QY 1129 TGTGAGAGGTGTTTGGGAAGTCCGAGCCCAACACATGATCTCCCGACCTCATCCAG 1188
Db 1457 TGTGTGAGATTGTTGGGCCCATGGGCAAGTACCAATGATGGCCCATCTTTCATCCAC 1516
QY 1189 ATCGACCTGTGCAACCCCTGTGACGCTGCTGATGATGATGATGATGATGATGATG 1248
Db 1517 GTGAACAAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1576
QY 1249 AGGGGACGCTGACTGCTCTGACCAACCCAGCAAGCCCATCTCCCTGCCAGAGAT 1308
Db 1577 GATGTCAGGAGATTGCTTCTGATGCCCCCACTCGGTTCTGCCCTCCCCACAGCC 1636
QY 1309 CTGCGGGGCCAG-----CTACACCTTGAGCCAGAGTGGAGTGGCTTTTGGCGTG 1362
Db 1637 CTCCCGGGCCACAGCACCTCTACGAGCTGGACAGCAGTGAACAGATCTTGGGCT 1696
QY 1363 GGCTCAACCCCTGCTTACAG-TGCACTGTGACCAAGCTGGTGGTGGTGGTGGTGGTGG 1421
Db 1697 GATTTCCGACACTGCCCCAACACCTCTGTGGAGGACATCTGTGACAGTCTGTGCCGT 1756
QY 1422 CAAGGACAGATGGTGTGCT-----AGACCCGCACTTCCCTGGGCCAT 1467
Db 1757 CATCGGATAGTATGAGGCCCATTTGCCACAAAGATGATGATGATGATGATGATGAT 1816
QY 1468 GGCACAGCTGTGGGAGGCAAGCTCTGCTCAAGGGGCTGGTGGGAGAGACACAA 1527
Db 1817 GGTACACCTGTGGGCTGGGACCTGTGCTGGATGATGATGATGATGATGATGATGATG 1876
QY 1528 CTCACAAAGCA-----GGGTGATGTTCTGGGCAATGGGATCCCTATGGCCCC 1581
Db 1877 GTGGAGATCCCAAGCTGTGTAGATGAGACTGGGTCCCTGGAGACCTGGGGGAA 1936
QY 1582 TGCTCGGCACATGTTGGGGGCTGACGCTGGCCAGGAGGAGTGCACCAACCCACC 1641
Db 1937 TGTTCTCGACCTGTGGTGGAGGATACAACTCTGCAACCTGATGATGATGATGATGATG 1996
QY 1642 CTGCGACAGGGGCAAGTACTGAGGAGTGGTGAATACCGATCTCTGCAACCTG 1701
Db 1997 CCTCAGATGGAGGAAGATTTCCTGGGTGAAGAGTCAAGTACCAATCATGCAACACA 2056
QY 1702 GAGCCCTGCCAGCTGACCTCCGGAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTC 1761
Db 2057 GAGGAATGTC-----CACCACCAAGGAAAAAGCTTCCGGGAGCAGCAGTGTGAGAAATAT 2110
QY 1762 AACGGCTACCAACACAGCACCAACCGGCTCACTCTCGCGGTGGCATGGTGGCCCAAGTAC 1821

Db 2111 AATGCCTACACCA---CACTGACCTGGATGGGAATTTCTGCACTGGGTCCCCAAGTAT 2167
QY 1822 TCCGCGTGTCTCCCGGGACAAAGTGAAGCTCATCTGCCAGCAATGCTACTGGGTAC 1881
Db 2168 TCAGAGTGTCCCCCGAGACCGATGCAAGCTGTTTTCAGAGCCCGTGGGAGG-NGTCA 2227
QY 1882 TTTATGTCCTGGACCCAAAGGTGGTGGACGGCAGCTGTCTCTGACTCCACCTCC 1941
Db 2228 TTCAAAGTGTGAAGCTTAAGTATGATCGATGGCCTCTGTGTGGACGGGATACTCTGTCC 2287
QY 1942 GTCTGTGCCAAGCAAGTGCATCAAGGCTGGCTGTGATGGGAACCTGGGCTCCCAAG 2001
Db 2288 ATCTGCGTCCGGGCAATGTTAAGGCTGCTGTGACCATGTTGTAACCTACCTAAG 2347
QY 2002 AGATTGCAAGTGTGGGCTGTGGGGGAGACAATAAGAGCTCCAAAGAGTGAAGTGA 2061
Db 2348 AAGCTGGAACAATGTGGGGTGTGGGGGCAAGGCACTGCTCTAGGAAGATCCTCGGT 2407
QY 2062 CTCCTTCAACCAAGCCATGATGGCTGCTACAATTTCTGTGGTGGCCATCCCGCAGGCGCTCA 2121
Db 2408 TCTTTCACCCCTTCAAGTTATGGCTACAATGACATTTGTCACCATCCAGTGGTSCACA 2467
QY 2122 AGCATCGACATCCCGCAGCGGTTTACAAGGGCTGATCGGGGATGACAATACTGGCT 2181
Db 2468 AACATTGATGTGAACAGCGGAGTCAACCGGGTCAAGGAGTCAAGGAGTGGTGGCG 2527
QY 2182 CTGAAGACAGCCAAAGCAAGTACTGCTCAACGGGCAATTTCTGTGGTGGTGGGCTGGAG 2241
Db 2528 CTGAAGACAGCCAAATGGGCAAGTACTGCTCAATGGTAACTGGCCATCTCTGCCATAGAG 2587
QY 2242 CGGACCTGTGTGTAAGGGCAGTCTGCTGGGTACAGCGGCAAGGCGGACAGCGGTGGAG 2301
Db 2588 CAAGCATCTGTGTGAAGGGACCATCTTGAAGTACAGTGGCTCCATGGCTACCTGGAG 2647
QY 2302 AGCTGAGGCTTCCCGGCCCATCTGGAGCGCTGACCGTGGAGGTCCTCTCCGTG--- 2358
Db 2648 CGGCTGAGAGCTTCCAGGGCCCTGCTGAGCCTCTTACAGTACAGCTCTGACTGTCT 2707
QY 2359 GGAAGATGACACCGCCCGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2408
Db 2708 GGTGAGTCTTCCCTCCAAAGTCAGATATACCTTTCTTGTGCCCAATGA 2757

RESULT 15

US-10-174-590-351

; Sequence 351, Application US/10174590

; Publication No. US20030008352A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3430RIC42

; CURRENT APPLICATION NUMBER: US/10/174,590

; CURRENT FILING DATE: 2002-06-18

; Prior application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 351

; LENGTH: 4407

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-174-590-351

Query Match

20.4%; Score 583.2; DB 9; Length 4407;

Best Local Similarity 60.4%; Pred. No. 4,1e-146;			
Matches 1093; Conservative 0; Mismatches 693; Indels 24; Gaps 57;			
QY	609	CCGGCGAGGTCTGGCGCGCCGAGCGTTTGGTGTATCTATCCCGCGGTAGCTGGAGACGCT	668
Db	1027	CAGCCCCAGACCCGAGAGCGCGTCTTCTTCACTGAGTAGATTTGTGGAGACACT	1086
QY	669	GGTGGTCCGGAGGAGTCAATGTCAAGTTCACGCGCGGACCTGGAACATTTATCTGCT	728
Db	1087	GGTGGTGGAGATGACAAGATGGCCGCAATTCACGGTGGGGCTTAAGCGCTACTGCT	1146
QY	729	GAGCGTCTGGCAACGCGCGGAGCTCTACCGCCATCCAGCATCTCAACCCCATCAA	788
Db	1147	AACAGTGTAGGACAGCAGAGCCCAAGCCCTCAAGCACCCCAAGCATCCGCAATCTGTCA	1206
QY	789	CATCGTTGGTCAAGGTGCTGCTTCTTAGAGATCTGACTCCGGCCCCAAGGTCAACGG	848
Db	1207	CTTGGTGGTCACTCGGCTAGTGATCTCTGGGGTTCAGCGAGGAGGCCCCAAGTGGGCC	1266
QY	849	CAATGGGGCCCTGACGCTGGCAACTCTTGCTGGCAGAGAAGCTGAACAAGTGAAG	908
Db	1267	CAGTGTGCCCGACCTGGCAGCTCTCTGCTGGCAGCGGGGCTCAACACCCCTGA	1326
QY	909	TGACAGACCCCGAGTACTGGGACACTGCCATCTCTTCAACAGCAGAGCTGTGTGG	968
Db	1327	GGACTCGGGCCCTGACCACTTTGACACAGCCATCTGTTTACCCTGTCAGGACCTGTGTGG	1386
QY	969	AGCCACACCTGTGACACCTGGGACGCTGATGTGGGTACCATGTGTGACCCCAAGAG	1028
Db	1387	AGTCTCACCTTGGACACGCTGGGTATGGCTGATGTGGGACCGCTGTGTGACCCGCTCG	1446
QY	1029	AAGTGTCTGTCTATTGAGGACGATGGGTTCCATCAGCTTCAACACTGCCACAGCT	1088
Db	1447	GAGCTGTGCCATTGTGGAGGATGATGGCTCCAGTCAAGCTTCACTGCTCATGAAT	1506
QY	1089	GGGCGAGTGTCAACATGCCCCATGACATGTGAAGTCTGTGAGGAGGTGTTGGGA	1148
Db	1507	GGGTCTATGTCTCAACATGTCTCATGACAACTCCAAGCCATGCATGTTGAATGGGC	1566
QY	1149	GCTCCGAGCCAA---CCACATGATGTCCCGACCCCTCATCCAGATCGACCGTCCCAACCC	1205
Db	1567	TTTGAGCACCTCTCGCATGTCTATGGCCCTGTGATGGCTCATGTGATCCTGAGGAGCC	1626
QY	1206	CTGCTAGCCTCGAGTGTCTCCATCATCATCGGACTTCTTGGACAGCGGGCAGGTAATG	1265
Db	1627	CTGCTCCCTCTGAGTCCCGCTTCTCATCTGACTTCTCTGGACAATGGCTATGGGCACTG	1686
QY	1266	CCTCTGGACCAACCCAGACGCCATCTCCCTGCGGAGGATCTCGCGGGCGCCAGCTA	1325
Db	1687	TCTCTTAGAACAAACAGAGGCTCCATGTGATCTGCTGTGACTTTCCTGGCAAGGACTA	1746
QY	1326	CACCTTGAGCGACGAGTGTGCTGCTTTTGGCTGGGTCCCAAGCCCTGTCTTACAT	1385
Db	1747	TGATGTGTACCGCCAGTGTGACCTCTGGGCGCGGACTCACGCCATTTGTCCACAGCT	1806
QY	1386	---CGAGTACTGACCAAGCTGTGGTCCACCGGAGGCCAAGGACAGATGTGTGCCA	1442
Db	1807	GCCCGCCCTGTGCTGCTCCCTCTGTGGCCACTCAATGGCCATGCCATGTGCCA	1866
QY	1443	GACCCGCACTTCCCTGGCGCGGATGACAGCTGTGGCGAGGCAAGCTCTGCCTCAA	1502
Db	1867	GACCAACACTCGCCCTTGGCGCGATGGCACACCTTGGCGCGCGCCAGCCCTGCATGGG	1926
QY	1503	AGGGGCTGTGGTGGAGACACACCTCAACAGCACAGGCT-----GGATGGTCTCTG	1556
Db	1927	TGCTGTGCTGCTCCACATGGACAGCTCCAGGACTTCAATATTTCCACAGGCTGTGGCTG	1986
QY	1557	GGCCAAATGGATCCCTATGGCCCTCTCTGCGCACATGTGTGGGGCGCTGCAGCTGGC	1616
Db	1987	GGGTCTTGGGGACCATGTGGGGTACTCTCTGGACCTGTGGGGGTGGTGTCCAGTTCTC	2046
QY	1617	CAGGAGGAGTGCACCAACCCACCCCTGCGCAACGGGGGCAAGTACTTGCAGGAGGTGAG	1676

Db	2047	CTCCCGAGACTGACAGAGGCGCTGTCCCCCGAATGTGGCAAGTACTGTAGGCGCCGCG	2106
QY	1677	GGTGAATACCATCTCTGCAACCTGGAGCCCTGCCAGCTCAGCCTCCGGAATAGAGCTT	1736
Db	2107	TACCGCTTCCGCTCTCTGCAACACTGAGGACTGCCCAA---CTGGCTCAGCCCTGACCTT	2163
QY	1737	CCGGAGAGGAGTGTGAGGCTTTCAAGGCTACAAACACAGCACCAACCGGCTCAGCTCT	1796
Db	2164	CCGGAGGAGGAGTGTGCTGCTTACAAACCCAGCCAGGACCTCTTCAAGAGCTTCCGAGG	2223
QY	1797	CGCCTGCATGGGTGCCCAAGTACTCCGGCTGTCTCCCGGGACAAGTGTCAAGCTCAT	1856
Db	2224	GCCCATGAGCTGGGTTCCTCTACAGGCGTGGCCCCCAGGACAGTGTGCAAACTCAC	2283
QY	1857	CTGCCAGGACCAATGGCACTGTGCTTCTATGTCTGTGACCCCAAGTGTGGAGCGAC	1916
Db	2284	CTGCCAGGCGCGGCACTGGGCTACTACTATGTCTGTGAGCCACGGTGTGATAGTGGAC	2343
QY	1917	GCTGTGCTCTCTGACTCCACTCGCTGTGTCTCAAGGCAAGTGCATCAAGGCTGGCTG	1976
Db	2344	CCCTGTCTCCCGGAGCTCTCTCGGTGTGTCTCAGGGCGGATGCATCCATGCTGGCTG	2403
QY	1977	TGATGGGAACCTGGGCTCCAAGAAGAGATTCCAAAGTGTGGGTGTGTGGGGAGACAA	2036
Db	2404	TGATCGCATCATTTGGCTCCAAGAGAAGTTTGAACAAGTGCATGTGTCCGAGGGGACGG	2463
QY	2037	TAAGAGCTGCAAGAAGTGTGACTGTGCTTCCAAAGCCCATGTGATGGCTTACAATTCGT	2096
Db	2464	TTCTGTTGACGACAGCAGTCAAGGCTCTCTCAGGAAATTCAGGTACGGATACAACAATGT	2523
QY	2097	GGTGGCATCCCGCAGGCGCTCAAGCATCGACATCCGCGCAGCGGTTTACAAGGCT	2156
Db	2524	GGTCACTATCCCGCGGGGCGCCACCATTTCTTGTCCGGCAGCAGGAAACCCCTGGCCA	2583
QY	2157	GATCGGGGATGACAACTACCTGGCTCTGAAGAACAAGCCAAAGTACCTGTCTCAAGCG	2216
Db	2584	CCGGAG-----CATCTACTTGGCCCTCAAGCTGCAGATGGCTCTATGCCCTCAATGG	2637
QY	2217	GCATTTGCTGTGTGCGGCTGGAGCGGGACCTGTGTGTG--NAGGGCAGTCTGC--TGCG	2273
Db	2638	TGAATACAGCTGTATGCCCTCCCCACAGATGTGTGTGCTGGGCGAGTCAJCTTGG	2697
QY	2274	GTACAGCGGCGGACAGCGGTGGAGAGCTGTGAGGCTTCCCGGGCCCATCTCTGGAGCC	2333
Db	2698	CTACAGCGGGCGCACTGTAGCCTCAGACACTGTTCAGGCCATGGCCAC--GGCCCAAGC	2757
QY	2334	GCTGACCTGTGAGGCTCTCTCCGTGGGAAAGATGACACCGCCCGGGTCCGCTACTCCTT	2393
Db	2758	TTTGACACTGCAAGTCTTAGTGGCTGGCAACCCCGAGGACACACGCTCCGATACAGCTT	2817
QY	2394	CTATCTGCC	2403
Db	2818	CTTCGTGCC	2827

Search completed: May 2, 2003, 00:40:34
Job time : 393 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 19:28:11 ; Search time 3890 Seconds
(without alignments)
11878.078 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgttttctggcgcaccc.....gcgtctcaggcgcgtctga 2853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:**
 - 2: em_esthum:**
 - 3: em_estin:**
 - 4: em_estmu:**
 - 5: em_estov:**
 - 6: em_estpl:**
 - 7: em_estro:**
 - 8: em_htc:**
 - 9: gb_est1:**
 - 10: gb_est2:**
 - 11: gb_htc:**
 - 12: gb_est3:**
 - 13: gb_est4:**
 - 14: gb_est5:**
 - 15: em_estfun:**
 - 16: em_estom:**
 - 17: gb_gss:**
 - 18: em_gss_hum:**
 - 19: em_gss_inv:**
 - 20: em_gss_pln:**
 - 21: em_gss_vrt:**
 - 22: em_gss_fun:**
 - 23: em_gss_mam:**
 - 24: em_gss_mus:**
 - 25: em_gss_other:**
 - 26: em_gss_pro:**
 - 27: em_gss_fod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	634	22.2	2931	11	BC003269 Mus muscu
2	531.8	18.6	849	13	BI103177 602889417
3	512.6	18.0	681	13	BG921734 602825569
4	406	14.2	961	10	BB612189 BB612189
5	391.4	13.7	522	10	BE553572 ur47b04.y
6	349.6	12.3	544	12	BF078689 229090 MA

7	339.4	11.9	627	10	BE281680
8	270.6	9.5	654	10	BE016461
9	263.4	9.2	738	13	BI819909
10	256.4	9.0	746	13	BI733795
11	245.8	8.6	661	10	BB220373
12	244.2	8.6	660	10	BB651049
13	242	8.5	662	10	BB642806
14	241.8	8.5	951	14	BQ930852
15	231.4	8.1	945	12	BF033017
16	229.4	8.0	575	13	BM570576
17	227.6	8.0	984	17	CNS0457M
18	226	7.9	797	14	BQ958765
19	222.6	7.8	305	12	BF320986
20	219.6	7.7	598	13	BM069815
21	215	7.5	674	10	BB225365
22	209.4	7.3	723	12	BG826611
23	209.2	7.3	949	14	BQ983675
24	206.2	7.2	652	9	A1400905
25	206	7.2	420	10	AW353152
26	205	7.2	681	12	BE912007
27	202.8	7.1	562	13	BJ009737
28	201.2	7.1	602	10	AW137467
29	200.2	7.0	601	14	BQ332355
30	196.4	6.9	680	12	BF446232
31	195.6	6.9	528	9	AA288689
32	190	6.7	518	17	AZ374992
33	188	6.6	689	12	BG242316
34	186.4	6.5	770	17	CNS03W7R
35	185.6	6.5	667	10	BG624639
36	185.4	6.5	504	12	BG553803
37	185.4	6.5	522	10	AW046509
38	184.8	6.5	640	12	BE742606
39	183.2	6.4	543	14	BQ011430
40	180.6	6.3	468	10	BE646878
41	180.6	6.3	475	9	A1148739
42	178.4	6.3	264	10	BB576723
43	177.4	6.2	566	10	AV617111
44	176.2	6.2	257	9	AI644268
45	172.4	6.0	464	10	BE666088

ALIGNMENTS

--- RESULT 1

BC003269 2931 bp mRNA linear HTC 07-AUG-2002

LOCUS Mus musculus, Similar to a disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 1, clone IMAGE:3499206, mRNA.

ACCESSION BC003269

VERSION BC003269

KEYWORDS HTC

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Strausberg, R.

AUTHORS 1 (bases 1 to 2931)

TITLE Direct Submission

JOURNAL Submitted (20-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk Email: cgabps@mail.nih.gov Tissue Procurement: Jeffrey Green M.D. cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: ang@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
 Toon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 9 Row: b Column: 7
 This clone has the following problem: frame shifted.

FEATURES

source

1. .2931
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="IMAGE:349206"
 /tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
 ductal carcinoma. 5 month old virgin mouse."
 /clone_lib="NCI_CGAP_Mam6"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 708 a 812 c 811 g 600 t

BASE COUNT

ORIGIN

Query Match 22.2%; Score 634; DB 11; Length 2931;
 Best Local Similarity 61.0%; Pred. No. 1e-125;
 Matches 1084; Conservative 0; Mismatches 680; Indels 12; Gaps 3;

Qy 631 AAGCGTTTCGCTATCCCGCGTAGTGGAGACGCTGGTGC GGGGACGAGTCAATG 690
 Db 1144 AAGCGATTGTGTCAGCCCGGTATGTGGAACCATGCTCGTGGTGACCACTCATG 1203
 Qy 691 GTCAAGTTCACGCGCGGAGCTTGAACATTTATCTGTGACGCTGTGCGCAACGCGGCG 750
 Db 1204 GCCGACTTCACGCGGAGGCTTAAAGCATTAACCTTAAACCTGTTCTCGGTGGCAGCC 1263
 Qy 751 CGACTTACCGCCATCCAGATCTCAACCCATCAACATCGTTGTGTCAGAGTCTG 810
 Db 1264 AGGTTTACAGCATCCAGCATAGGAATTCATTAAGCTGTGTGGTGGTGAAGATCTTG 1323
 Qy 811 CTCTCTAGAGATCTGACTCGCGGCCCAAGGTCAACGGCAATCGGCGCTGAGCGTCCG 870
 Db 1324 GTCATATAGGAGAGAGAGGACCAAGATTAACCTTCAACCTGATCGAGCTTCAACCTTCGG 1383
 Qy 871 AACTTCTGTGCTGGCAGAGAAGCTGAACAAAGTGTAGTGAACAGCACCCGAGTACTGG 930
 Db 1384 AATTCTCAACTGGCAGAAACACACACAGCCCAAGTACCGGGATCCAGAGCACTAT 1443
 Qy 931 GACACTGCCATCTTTCACGAGCAGGACCTGTGTGAGCCACCACCTGTGACACCTCG 990
 Db 1444 GACACTGCAATTTCTGTTACACAGACAGGATTTATGTGGCTCCACACAGTGTGACACTCTC 1503
 Qy 991 GSCATGGCTGATGGGTACCATGTGTACCCCAAGAGAAGCTCTCTGTCAATTGAGGAC 1050
 Db 1504 GGGATGGCAGATGTGGAACTGTATGTGACCCAGCAGGAGTCTCAGTCATAGAAGAT 1563
 Qy 1051 GATGGGCTTCCATCAGCTTCCACACTGCCCCAGAGTGGGCCACGTTGTCAACATGCC 1110
 Db 1564 GATGGTTTGAAGCGCGCTTCCACACAGCCCAAGATTTGGGCCATGTGTTTAACATCCG 1623
 Qy 1111 CATGACATGTGAAAGTGTGTGAGGAGTGTGGAAGCTTCCGAGCCCAACCATCATG 1170
 Db 1624 CACGATGATGTAAAGCATGTGCGCAGCTTGAATGGTGTGACTGGCGATTTCTATCTGATG 1683
 Qy 1171 TCCCGGACCTTCATCCAGATCGAGCTGCCAACCCCTGGTCAGCCTGCAGTGTGCCATC 1230
 Db 1684 GCCTGATGCTTCCAGCTTAGACCATAGCCAGCCCTGGTCACCTTGCAGTGCCTACATG 1743
 Qy 1231 ATCACCGACTTCTTGACAGCGGCGAGGTGACTGCTCTCTGGACCAACCCAGCAAGCC 1290
 Db 1744 GTCAGCTCTTCTAGATAATAGGACACGGGAATGTTTGTATGGACAAAGCCCAAGATCCA 1803
 Qy 1291 ATCTCTCTGCCGAGGATCTCGCGGGCCAGACTACACCTTGAGCCAGCTGCGAGCTG 1350

Db 1804 ATCAAGCTCCCTTCTGTATCTTCCCGGTACCTTGTACGATGCCAACCCGCACTGTGAGTTT 1863
 Qy 1351 GCTTTTGGCGTGGGCTCCAAAGCCTGTCTTACATGCAGTAC---TGACCAAGAGTTGTGG 1407
 Db 1864 ACATTGGAGAGGAATCCACGACATGCCCTGATGCGCCAGCAGCATGTACTACTCTGTGG 1923
 Qy 1408 TGCAACGGGAAGGCCAAGGAGCAGATGGTGTGCCAGAGCCGCCACTTCCC"TGCGCCGAT 1467
 Db 1924 TGCACTGGCAGCTCGGTTGGTGGTACTTGTGTGTCGCAAAACAAACACTTCCCCTGGCAGAT 1983
 Qy 1468 GGCACAGCTGTGGGAGGGAAGCTCTGCCTCAAGGGGCTGCGTGGAGAGACACACAC 1527
 Db 1984 GGCACAGCTGTGGAGAGGAAGTGGTGTGTAGTGGCAAGTGGTGTGCAACAGACAGAC 2043
 Qy 1528 CTCACCAAGCAGCAG-----GGTGTGATGTTCTTGGGCGCAATGGGATCCCTATGGCCCC 1581
 Db 2044 ATGAAGCATTTTGTCTACTCTCTTCTATGGAAGCTGGGACCATGGGACCGTGGGAGAC 2103
 Qy 1582 TGCTCGCGACATGTGGTGGGCGGTGAGCTGGCCAGGAGGAGTGCACCAACCCACC 1641
 Db 2104 TGCTCAAGAACCTGTGGTGGTGGAGTTCAATACACAAATGAGAGAAATGTGACACCCAGTC 2163
 Qy 1642 CCTGCCACGGGGGCAAGTACTGCGAGGAGTGAAGGTGAATACCGATCCTGCAACCTG 1701
 Db 2164 CCAAGAGACGGAGGAGTACTGTGAAGCAACAGTCCGCTACAGTCTCTGTAAACATC 2223
 Qy 1702 GAGCCCTGCCCCAGCTCAGCTCCCGAAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTC 1761
 Db 2224 GAGGACTGTCCAGACAATAA---CGGAAAAAGCTTCAGAGAGGAGCAGTGTGAGGCGCAC 2280
 Qy 1762 AACGGCTACCAACCAAGCAGCCAGCTCTCGCGGTGGCATGGTGGTGGTGGTGGTGGTGGTGG 1821
 Db 2281 AATGAGTTTTCACAAAGCTTCTTTGGGAATGAGCCACTGTAGAGTGGACACCCAGTAC 2340
 Qy 1822 TCCGGGTGTCTCCCGGGAGCAAGTGCAGAGTCACTTGCAGAGCCAAATGGCAGTGGCTAC 1881
 Db 2341 GCGGGGTCTCCGCAAGGACAGTGCAGCTCACTGTGAAGCCCAAGGATTTGGCTAC 2400
 Qy 1882 TTCTATGTCTGGCACCCAAAGTGGTGGAGCGGAGCTGTGCTCTCTCTGACTCCACCTCC 1941
 Db 2401 TTTTTCGTCTTACAGCCCAAGGTGTAGATGGCAGCTCCCTGTAGTCCAGACTCTACCTCT 2460
 Qy 1942 GTCTGTCTCAAGGCAAGTGCATCAAGCTGTGTGTGAGGAACTGGGCTCCCAAGAG 2001
 Db 2461 GTCTGTGTGCAAGGCGAGTGTGAAGCTGTGTGAAGCTGTGTGATCGCATATAGACTTCCAAAG 2520
 Qy 2002 AGATTGCAAGTGTGGGTGTGTGGGGAGACAATAAGAGCTGCAAGAGTGAAGTGAAGTGA 2061
 Db 2521 AGTTTGTAGTGTGGCGTTTGTGGAGGAAACGGTTCCACATGCAAGAAGATGTCAAG 2580
 Qy 2062 CTCCTTCCACCAAGCCATGATGGCTACAAATTTCTGGTGGCCATCCCGCAGCGCCCTCA 2121
 Db 2581 ATAGTCACTAGTACAAAGCTGGGTATCATGACATTTGTCAAAATTCCTGTCTGGAGCCAC 2640
 Qy 2122 AGCATCGACATCCGCGAGCGGTTTACAAAGGCTGATCGGGGATGACAACTACCTGGGT 2181
 Db 2641 AACATTGAAGTGAACATCGGAATCAAGGGGTCCAGAAACAAATGGCAGCTTCTGGCT 2700
 Qy 2182 CTGAAGAACAGCAAGGCAAGTACCTGTCAACGGGCAATTCGTGGTGTGGCGGTGGAG 2241
 Db 2701 ATTAGAGCCGCTGATGGTACCTATATCTGAATGGAACCTTCACTCTGTCCACACTAG 2760
 Qy 2242 CGGGACCTGGTGTGAAGGCGAGTCTGCTGCGGTACAGCGGCGCAGCGGACAGCGGTGGAG 2301
 Db 2761 CAAGAGCTTCACTTACAAAGGTACTGTCTTAAGGTACAGTGGTCTCGGCTCGGCTGGAG 2820
 Qy 2302 AGCCTTCAGGCTTCCCGGCGCATCTCTGGAGCGGCTGACCGTGGAGGCTCTCTCCGTGGGG 2361
 Db 2821 AGAATCCGAGCTTTAGTCCACTCAAGAAACCTTAACCATCCAGGTTCTTATGTGTAGGC 2880
 Qy 2362 AAGATGACACCGCGCGGTGCGCTACTCTCTCTAT 2397

```

Db 2881 CATGCTCCGACCAAAATTAATTCACCTACTTT 2916
      BII03177      849 bp      mRNA      linear      EST 26-JUN-2001
      LOCUS      602889417F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5044493
      DEFINITION      5', mRNA sequence.
      ACCESSION      BII03177
      VERSION      BII03177.1 GI:14554070
      KEYWORDS      EST.
      SOURCE      house mouse.
      ORGANISM      Mus musculus
      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
      1 (bases 1 to 849)
      NIH-MGC http://mgc.nci.nih.gov/
      National Institutes of Health, Mammalian Gene Collection (MGC)
      Unpublished (1999)
      Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-r@mail.nih.gov
      Tissue Procurement: Jeffrey E. Green, M.D.
      CDNA Library Preparation: Life Technologies, Inc.
      CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: L1AM1122 row: a column: 06
      High quality sequence stop: 743.
      Location/Qualifiers
        1..849
          /organism="Mus musculus"
          /strain="FVB/N"
          /db_xref="taxon:10090"
          /clone="IMAGE:5044493"
          /clone_lib="NCI_CGAP_Kid14"
          /lab_host="PH10B (T1 phage-resistant)"
          /note="Organ: kidney; Vector: pCMV-SPORT6; Site.1: NotI;
          Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
          Average insert size 1.75 kb. Constructed by Life
          Technologies. Note: this is a NCI_CGAP Library. |"
      BASE COUNT      193 a 250 c 250 g 156 t
      ORIGIN
      Query Match      18.6%; Score 531.8; DB 13; Length 849;
      Best Local Similarity 85.1%; Pred. No. 6.6e-104;
      Matches 699; Conservative 0; Mismatches 112; Indels 10; Gaps 9;

QY 1189 ATGACCGTGCACACCC-TGGTCAGCGTGCAGTGTGCCATCATCAC-CGACTTCTCTGG 1246
      |||||||
Db 26 ATAGACCGTGCCACCCCATGGTCAGACTGCAGCGCTGCATTATAACGTGACTTCTCTGG 85
      |||||||

QY 1247 ACAGCGGCGACGTGACTGCTCTCGACACACCCAGCAGCCATCTCCCTGCCGAGG 1306
      |||||||
Db 86 ATAGTGGCATGTGACTGCTCTCGACACCCAGCAGCCATCATCTCTGCTGCTGAGG 145
      |||||||

QY 1307 ATCTGCGGCGCGACGTACACCTGAGCCAGTGCAGTGTGGCTTTGGCTGGGCT 1366
      |||||||
Db 146 ACCTGCGGCGCACAGCTACAGTTTCAGCCAGCAGTGTGGCTTTGGGTTGGGCT 205
      |||||||

QY 1367 CCAAGCCCTGCTTACATGCACTACTGCACCAAGTGTGGTGACCCGGAAGCCAAAG 1426
      |||||||
Db 206 CTAAGCCCTGCCCATATATGCACTACTGTACAAGCTGTGGTGCAAGGCAAGGCAAG 265
      |||||||

QY 1427 GACAGATGTTGCCAGACCGGCACTTCCCTGGCCCATGGCACCAGCTGTGGCAGG 1486
      |||||||
Db 266 GGCAGATGTTGCCAGATCGCCACTTCCCTGGGCAATGGCACCAGCTGTGGGGAAG 325
      |||||||

QY 1487 GCAAGCTCTGCTCAAGGGGCTGGTGGAGAGACACACCTCAACAGCAGCAGGTTGG 1546
      |||||||
Db 326 GCAAGTTCTGCTCAAGGGAGCTGGTGGAGAGACACACCAACCAACAGTACCGGTTGG 385
      |||||||

QY 1547 ATGTTCTTGGGCGCAAAATGGATCCCTATGGCCCTCTGTCGCGACATGTGTGGGGGG 1606
      |||||||
Db 386 ACGGCTCTTGGGCGCAAGTGGGAGCCCTACGGTTCTGCTTCCGCTGCGGTGGGGGG 445
      |||||||

QY 1607 TGCAGCTGCGCAGGAGGAGTGCACCAACCCACACCC-TGCCAACGGGGGGLAGTACTGC 1665
      |||||||
Db 446 TGCAGCTGCGCAGGAGGAGTGCAGCAACCCCTACCCCATGCCACAGGTGGAAAATACTGC 505
      |||||||

QY 1666 GAGGAGTGAAGGTGAATACCGATCTCTCAACCTGGAGCCCTGCCAGCTCAGCCTCC 1725
      |||||||
Db 506 GAGGAGTGAAGGTGAATACCGATCTCTCAACCTGGAGCCCTGCCAGCTCAGCCTCT 565
      |||||||

QY 1726 GGAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCACCAC 1785
      |||||||
Db 566 GGAAGAGCTTCCGGGAGGAGCAATGTGAAGCTTTCAATGGCTACAAAGCAGACAGAAC 625
      |||||||

QY 1786 CGGCTCAGCTCTCGCCGTGGCATGGTGCCTCAAGTACTCTCGGCGTGTCTCCCGGAGCAAG 1845
      |||||||
Db 626 CGGCTCAGCTTACGTGTGCATGGGTACCCAAAGTACTCAGCGGTGTACAGCTGACCAAG 685
      |||||||

QY 1846 TGAAGCTCATCTGCGGAGCCAAATGGCACTGGCTACTTCTTATGTGCT-GGCACCCAAAGT 1904
      |||||||
Db 686 TGAAGCTCATCTGCGGAGCC-ATGGGACTGGCTACTTCTTATGTACTAGACACCTAAGT 744
      |||||||

QY 1905 GTTGA-CGCGACGCTGTG-CTCTCTGACTCCACTCCGCTGTCTTCAAGGGCAAG-TG 1961
      |||||||
Db 745 GTTGA-CGCGACGCTGTG-TAAGCTGTAACACTGACTCCACTCGGTGTGTCTCAAGGCAAGT 804
      |||||||

QY 1962 CATCAA--GGTGTGCTGTGATGGGAACCTGGGCTCCAAGAA 2000
      |||||||
Db 805 CATCCACGGGCTGTGTCGAAGCGCAACCTGGGCTCCAAGCA 845
      |||||||

RESULT 3
BG921734
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
  1..681
    /organism="Mus musculus"
    /strain="FVB/N"
    /db_xref="taxon:10090"
    /clone="IMAGE:4954456"
    /clone_lib="NCI_CGAP_Mam6"
    /sex="female, virgin"
    /tissue_type="infiltrating ductal carcinoma"
    /dev_stage="5 months"
    /lab_host="DH10B"
    /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
    Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT."

```


Query Match 14.2%; Score 406; DB 10; Length 961;
Best Local Similarity 82.7%; Pident No. 7.3e-77;
Matches 488; Conservative 0; Mismatches 100; Indels 2; Gaps 2;

Qy 1 ATGCTTTTTCGTGGGCATCCTAACCCCTGGCTTTTCGCCGGCGCAACGCTGGAGGCTTTTGAG 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 321 ATGCTTCTACTTGGGCATCTCAATCCTGGCTCTGCCTGGCGACCTGCTGGCAGCTCCGAG 380
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 61 CCAGAGCGGGAGGTAGTTCGTTTCCCATCCGACTGGACCCGGACATTAACGGCCGCGCTTAC 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 381 CCAGAGTGGGAGGTGGTTCGTTTCCAAATCCGAGCGGACCCGGACATCAATGGCGCCACTAC 440
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 121 TACTTGGCGGGTCCCGAGGACTCCGGGGATCAGGAGCTCATTTTTTCAGATCACAGCATTT 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 441 TACAGGAGGGTAGGAGGACTCCGGGGATCAGGCTCATTTTTTCAGATCACAGCATTT 500
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 181 CAGGAGGACTTTTACCCTACACTGACGCCGATGCTCAGTTCTTTGGCTCCGCGCTTCTTCC 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 501 CAGCAGGACTTTTATCTACACTGACGCAGATGCCAGTTCTTGGCTCCGCGCTTCTTCT 560
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 241 ACTGAGCATCTGGGGTCCCGCTCCAGGGGCTCACC GGCGCTCTTACAGACTGGGAGCG 300
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 561 ACTGAGTATCTAGGTGTCCTCCGCTCAGAGGCTCACTGGCAGCTCTCTAGACTCGCAGCG 620
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 301 TGCTTCTATTCTGGGAGCTGAAGCCGAGCGGACTCGTTGCTGCTGTGAGCCTGTGC 360
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 621 TGCTTCTATTCCGGTAGTGAACGGGAG -CAGACTCTTTGCTGTGAGCCTATGC 679
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 361 GGGGGCTCCCGGAGCCTTTTGGCTACCGAGCGCGGAGTAGTATGATTTAGCCCGCTGCC 420
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 680 GGGGGCTCCCGGAGCCTTTTGGCTACCGAGGTCGAGAGTATGATTTAGCCCTCTGCC 739
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 421 AATGTAGCGCGCGCGGCGCAGCGAACAAGCAGGGCGCACACTTCTCCAGCGCGG 480
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 740 AACACGCGCGCCAGAGCGAGCGTGACAGCAGCGGCGCACACTTCTCCAGCGCGG 799
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 481 GGTGTTCCGGCGGGCTTCCGGAGACCCCACCTCTCGCTCGGGGGTGGCTCGGGGTGG 540
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 800 GGTGCTCTGTATGGCTTCCGGAGACACACCTCTCGTTGCGGGGTGG -CTCGGGCTGG 858
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 541 AACCCGCGCATCTACGGGCGCTGACCCCTTACAAGCGCGCGCGCGGCGG 590
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 859 TACCGGCATCCCTGAGGCTCTCGACCCCAAGCCACCGGAGGCGCGG 908
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5 BE553572 522 bp mRNA linear EST 15-AUG-2000
LOCUS ur47B04.y1 NCI-CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153391 5'
DEFINITION similar to TR:O54768 O54768 A DISINTEGRIN AND METALLOPROTEASE WITH THROMBOSPONDIN MOTIFS ; , mRNA sequence.

ACCESSION BE553572
VERSION BE553572.1 GI:9817989
KEYWORDS EST.
SOURCE house mouse,
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus. 1 (bases 1 to 522)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: craps@remail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1056147
Seq primer: -40rp from Gibco
High quality sequence stop: 385.
Location/Qualifiers 1. .522
/organism="Mus musculus"

COMMENT

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@meat.marc.usda.gov

Single: pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCA
BACKWARD: GTTTCCTCAGTCAGCAGG
Plate: 43 row: P column: 17
Seq primer: ATTAGGTACACTATAG.

FEATURES

source

1..544
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site.1: NotI; Site.2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 117 a 154 c 169 g 104 t

Query Match 12.3%; Score 349.6; DB 12; Length 544;

Best Local Similarity 88.6%; Pred. No. 8.1e-65; Mismatches 49; Indels 0; Gaps 0;

559 GCCTGACGCTTACAAAGCGCGGGGGCTTCGGGAGAGTGTAGCGGGCGCAGG 618

544 GCCTGGACGCTTACAAAGCTTCGGGATAGGCTTAGGGAGGTCGACGCGCGCAGG 485

619 TCTGGGCGGCCAAGCGTTTCGTGTCTATCCCGGTAGCTGGAGACGCTGTGTGCGG 678

484 TCGGGGCGGCCAAGCGCTTCGTGTCTATCCCGGTAGCTGGAGACGCTGTGTGCGA 425

679 GAGAGTCAATGTCAAGTTCACAGCGCGGACCTCGCAACATTATCTGTGACGCTGTG 738

424 GACAGTCCATGTCAAGTTCACAGCGCGGACTTGGAGCATTATCTGTGACGCTGTG 365

739 GCAACGGCGCGGACTCTACCGCCATCCAGCATCCTCAACCCCATCAACATCGTTGTG 798

364 GCCACGGCGCGGACTCTACCGCCATCCAGCATCCTCAACCCCATCAACATCGTCGTG 305

799 GTCAAGGTGCTTCTTAGAGATCTGACTCCGGGCCCAAGGTCAACGGCAATCGGCCC 858

304 GTCAAGGTGCTTCTTAGAGACCGCGACACAGGCCCAAGGTCAACGGCAACCGGCC 245

859 CTGACGCTGGCAACTTCTGTGCTGGCAGAGAGCTGAACAAAGTGAAGTGAACAGCAC 918

244 TTGACGCTGGCAACTTCTGTGCTGGCAGAGAGCTGAACAAAGTGAAGTGAACAGCAT 185

919 CCGGAGTACTGGGACACTGCCATCTCTTTCACAGCAGGACCTGTGTGGAGCCACAC 978

184 CTTGAGTACTGGGACACCGCATCTCTTTCACAGCAGGAGTGAAGTGAAGTGAAGTGA 125

979 TGTGACAC 986

124 TTGAGCCC 117

RESULT 7

LOCUS

BE281680

DEFINITION

601099657F1 NC1_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3491991 5',

Accession

BE281680

Version

BE281680.1

Keywords

EST.

Source

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 627)

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: coapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8537 row: a column: 16
High quality sequence stop: 530.

FEATURES

source

1..627
Location/Qualifiers

/organism="Mus musculus"

/strain="Czech II"

/db_xref="taxon:10090"

/clone_image="3491991"

/clone_lib="NCI_CGAP_Lu29"

/tissue_type="spontaneous tumor, metastatic to mammary."

Stem cell origin."

/lab_host="DH10B"

/note="Organ: lung; Vector: PCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 153 a 159 c 177 g 138 t

ORIGIN

Query Match 11.9%; Score 339.4; DB 10; Length 627;

Best Local Similarity 83.6%; Pred. No. 1.3e-62;

Matches 504; Conservative 0; Mismatches 86; Indels 13; Gaps 10;

QY 1850 AGCTCATCTGCGGAGCAATGGCCTACTTCTATGTG-CTGGCACCACCAAGTGGTG 1908

DB 1 AGCTAATCTGCGGAGCAATGGGACTGGCTACTTCTATGTAAGTACACCTAAGCTGGTG 60

QY 1909 GACGACAGCTGTGCTTCCTGAC-TCCACCTCGTGTGTGTCCCAAGCAAGTGCATCAA 1967

DB 61 GATGTACGCTGTGTACTCTCTGACATCCACCTCGTGTGTGTCCCAAGCAAGTGCATCAA 120

QY 1968 -GGCTGGCTGTGTTGGGAACTGG-GCTCCAAAGAGATTTCGACAA-GTGTGGGGTGTG 2024

DB 121 TGGCTGGCTGGCAGGGAATCTGGTCTCCAAAGAGAAATTTGACACATGTGTGTGTG 180

QY 2025 TGGGAGACACAATAA-GAGCTGCAAGAGGTGACTGGACTCTTCCCAAGCCCAATGCATG 2083

DB 181 TGGTGGACACAATAATGAGCTGTAAGAGGTGACAGGACTCTTCCCAAGCCCTTGCACG 240

QY 2084 GCTACAATTCGTGTGGCCATCCCGCAGCGGCTCAAGCATGCATCGGCAGCGCG 2143

DB 241 GCTACAATTCGTGTGGCCATCCCGTGTGGCCCTCCAGCATGTATATCGTCAGCGTG 300

QY 2144 GTTCAAGAGGCTGATCGGGATGACACTACCTGGCTCTGAAGAACAGCCCAAGGCAAGT 2203

DB 301 GTTCAAGAGGCTCATTTGGAGATGACAACTACCTGGCTTTGAAGAACAGCCCAAGCAAT 360

QY 2204 ACCTGCTCAAGGGCATTTTCGTGTGTGTCGGCGGTGGAGCGGACCTGGTGGTGA--GGG 2261

DB 361 ACCTGCTCAATGGGCACTTACTGTATCCGCTGTAGAGCGGACCTGTGTGTAAACAGG 420

QY 2262 CAGTGTGTCGGGTACACGGCAGCGGACAGCGGTGGAGAGCTGCAGGCTTCCCGGCC 2321

DB 421 CAGTGTGTCAGCTATAGTGGCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 479

QY 2322 CATCTGTCAGCGCTGACCGTGGAGGTCTCTCCGTGGGGAAGATGACACCGCCCGCGGT 2381


```
/clone="IMAGE:5176476"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/Note="organ: pooled brain, lung, testis: Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dr primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH MGC Library."

BASE COUNT      169 a   204 c   225 g   140 t
ORIGIN
Query Match      9.2%; Score 263.4; DB 13; Length 738;
Best Local Similarity 63.1%; Pred. No. 2.9e-46;
Matches 440; Conservative 0; Mismatches 251; Indels 6; Gaps 2;

QY 1715 GCTCAGCCTCCGGAAGAGCTTCGGGAGGACAGTGTGAGGCTTTCACGGCTACACACC 1774
DB 10 GCGCCCTCGCGGAAAGCTTCAGGAGCAGAGTGTGAGAGTATATGCTACAAAT 69
QY 1775 ACAGCACCAACCGCTCACTTCGCGGTGGCATGGTGCCCAAGTACTCCGGCGTGTCTC 1834
DB 70 ACACAGTACAGGACGGAATCT---CCTGAGTGGGTCTCCCAAGTATGCTGGGTGTCC 126
QY 1835 CCGGGACAAAGTCAAGTCTATGCGGAGCCATGGCACTGGCTACTTCTATGTGCTGG 1894
DB 127 CCGGGACCGCTCAAGTGTGTGCGGAGCCCGGGAGGAGCGAGTTCAAAGTGTTCG 186
QY 1895 CACCAAGGTGTGGACGACCTGTCTCTCTCACTCCACCTCCGCTGTGTGCCAAG 1954
DB 187 AGCCAAAGTGTATGATGCAACCTGTGTGGCCCAACAACTGGCCACTGTGTCCGTG 246
QY 1955 GCAAGTGCATCAAGGTGGCTGTGTATGGAACTGGGCTCCCAAGAGAGATTGCAAT 2014
DB 247 GCCAGTGTGTCAAGGCGCGGTGTGACCATGTGTGAGCTCGCCTCGGAAGCTGGACAA 306
QY 2015 GTGGGTGTGTGGGAGACAAATAGAGCTGCAAGAGGTGACTGGACTTTCACCAAGC 2074
DB 307 CGGGGTGTGTGGGGCAAGGCAACTCTGCGAGGAAGTCTCCGGGTCCCTCACCCCA 366
QY 2075 CCATGATGCTACAATTTGTTGTTGGCCATCCCGAGCGGCTCAAGCATCGACATCC 2134
DB 367 CCAATTATGCTACAATGACATTTGCACATCCCACTGGTGCCACTAATATTGACGTGA 426
QY 2135 GCACGCCGGTTCAAAGGCGCTATCGGGATGACAACTACCTGGCTCTGAAGACAGCC 2194
DB 427 AGCAGCGGACCCCGGGGTGTGCAGAACGATGGGAACCTACCTGGCGCTGAAGCGGTG 486
QY 2195 AAGCAAGTACCTGCTCAACGGCATTTGCTGTGTGCGGGTGGAGCGGACCTGGTGG 2254
DB 487 ATGGCGAGTACCTGCTCAACGGCACTTGCCATCTCTGCCATAGACGAGCATTTGG 546
QY 2255 TGAAGGCACTGTGCTGCGGTACAGCGGCACGGGACAGCGGTGGAGAGCTTCAGGCTT 2314
DB 547 TGAAGGGGACCATCTCAAGTACAGCGGCTCCATCGACACCTTCGAGCGGCTCGAGACT 606
QY 2315 CCGGGCCATCTCGAGCGCTGACCGTGGAGGTCTCTCTCCGT---GGGGAAGTACAC 2371
DB 607 TCGGGCCCTTGGCAGAGCCCTGTGACAGTGCAGCTCTCTGACAGCTCCCTGGGAGGTTC 666
QY 2372 CGCCCGGGTCCGCTACTCTCTTCTATCTGCCCAAGA 2408
DB 667 CCCAAAAGTCAATACACCTTCTTGTACCTAATGA 703

RESULT 10
BI733795      746 bp      mRNA      linear      EST 20-SEP-2001
LOCUS        BI733795
DEFINITION   603351940F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5359712 5',
```

```
mRNA sequence.
BI733795
VERSION      BI733795.1 GI:15710808
KEYWORDS
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 746)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE        Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
COMMENT      Unpublished (1999)
              National Institutes of Health, Mammalian Gene Collection (MGC)
              Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-r@mail.nih.gov
              Tissue Procurement: The Cepko Laboratory
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM1914 row: 0 column: 09
              High quality sequence stop: 746.
              Location/Qualifiers
                source          1..746
                organism="Mus musculus"
                db_xref="taxon:10090"
                clone="IMAGE:5359712"
                clone_lib="NIH_MGC_94"
                tissue_type="retina"
                lab_host="DH10B (phage-resistant)"
                note="organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
                Average insert size 3.3 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."

BASE COUNT      213 a   160 c   216 g   156 t   1 others
ORIGIN
Query Match      9.0%; Score 256.4; DB 13; Length 746;
Best Local Similarity 61.5%; Pred. No. 9.4e-45;
Matches 448; Conservative 0; Mismatches 272; Indels 9; Gaps 2;

QY 1463 CCGATGGCACCAGCTGTGCGGAGGCAAGCTCTCCCTCAAGGGGCGCTCGTGGAGAGAC 1522
DB 1 CAGATGGCACCAGCTGTGGAAGGGAAGTGTGTGTGTCAGTGGCAAGTGTGCTGACACAGA 60
QY 1523 ACAACCTCAACAA-----GCACAGGGTGGATGTCTCTGGGCCCAATGGGATCCCTATG 1576
DB 61 CAGACATGAGCATTTTCTACTCTCTCATGGAAGCTGGGGACCATGGGACCGTGGG 120
QY 1577 GCGCCCTGTGCGGCACATGTGTGGGGCGGTGCAGCTGGCCAGGAGGAGTGCACCAACC 1636
DB 121 GAGACTGTCTCAGAAACCTGTGTGTGGAGTTCATACACAATGAGAAATGTGACAAACC 180
QY 1637 CCACCCCTGCCAACGGGGCAAGTACTCGAGGAGTGCAGGGTGAATACGATCCCTGCA 1696
DB 181 CAGTCCCAAGAACGGAGGAGTACTGTGAAGGCAACAGAGTCCGCTACAGCTCTGTA 240
QY 1697 ACCTGGAGCCCTGCCCCAGCTCAGCTCCGGAAGAGGTTCGGGAGGAGGAGTGTGAGG 1756
DB 241 ACATCGAGGACTGTCCAGAC---AATAACGGAAAAACCTTCAGAGGAGGAGTGCAGG 297
QY 1757 CTTTCAAGGGCTACAAACACACACCAACCGGCTCACTTCGCGTGCATGGGTGGCCCA 1816
DB 298 CGCACAATGAGTTTCCAAAGCTTCTNTGGGAATGAGCCCACTGTAGTGGACACCCA 357
QY 1817 AGTACTCCGGCGTGTCTCCCGGGGACAAAGTCAAGCTCATCTGCGGAGGCCAATGCACTG 1876
DB 358 AGTACGCGCGCTCTCGCCAAAGGACAGTGCACACTCACCTGTGTGAAGCCAAAGCAATTG 417
QY 1877 GCTACTTCTATGTGTGGACCCCAAGGTGGTGGACGCGTGTGCTCTCTCTGACTCCA 1936
```

Db	418	CTACTCTTTTGGCTTTACAGCCCAAGGTGTAGATGGCACTCCCTGTAGTCCAGACTCTA	477
QY	1937	CTCCGCTGTGTCTTCAAGGCAAGTGCATCAAGCGTGTGTGATGGGAACCTCGGGCTCCA	1996
Db	478	CTCTCTGTCTGTGCAAGGGCAGTGTGTGAAAGCTGGCTGTGATCGCATATAGACTCCA	537
QY	1997	AGAAGAGATTCCACAAAGTGTGGGTGTGTGGGGGAGACAATAAGAGCTGCAAGAAGGTGA	2056
Db	538	AAAAGAAGTTGGATAGTGTGGCGTTTGTGGAGAGAAACGGTTCACATGCAAGAAGATGT	597
QY	2057	CTGGACTCTTCCACCAAGCCCATGCATGGCTACAAATTTCTGGTGGCCATCCCGCAGGCG	2116
Db	598	CAGGAATAGTCACTAGTACACAGACTGGGTATCATGACATGGTCAATTCCTGTCTGGAG	657
QY	2117	CCTCAAGCATGCACATCCGCGCAGCGCGGTGTACAAAGCGGTGATCGGGGATGACAACTACC	2176
Db	658	CCACCACAACTGAAGTGAACATCGGAATCAAAAGGGGTCCAGAAACAATGCGAGCTTTC	717
QY	2177	TGGCTCTGA	2185
Db	718	TGGCTATTA	726
RESULT	11		
LOCUS	BB220373		
DEFINITION	BB220373 RIKEN full-length enriched, adult male aorta and vein Mus musculus CDNA clone A33059019 3' similar to AF140673 Mus musculus putative secreted metalloprotease ADAMTS5 (Adamts5) mRNA, mRNA sequence.		
ACCESSION	BB220373		
VERSION	BB220373.2	GI:16353495	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.		
TITLE	RIKEN Mouse ESTs (Arakawa, T., et al. 2001)		
JOURNAL	UNPUBLISHED (2001)		
COMMENT	On Jun 30, 2000 this sequence version replaced gi:8885326. Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.riken.go.jp URL: http://genome.gsc.riken.go.jp/ Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a		


```

QY 1318 GCCAGCTACACCTGACGCCAGCAGCTGCGCTGCTTTGGGTGGCTCCCAAGCCCTGT 1377
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 545 CAGACCTACCATGCCACCCAGCAGTGCACCTTGACATTTGGGCTGAGTCTCGGTGTC 604
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1378 CTTTACCTAGTACTGTCACCAAGCTGGTGCACCGGGAAGCCCAAGG 1426
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 605 CTTGGCATGATCTGTGCGGGCTGTGTGTGCTGTGTGCGGCAGG 653
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 13
BB642806
LOCUS BB642806 RIKEN full-length enriched, adult retina Mus musculus cDNA
DEFINITION clone A930039D02 5', mRNA sequence.
ACCESSION BB642806
VERSION BB642806
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 662)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
UNpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
Ishii,Y. and Hayashizaki,Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
Func. Genomics 2 pre, L72-L86 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
Location/Qualifiers
1. 662
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A930039D02"
/clone_lib="RIKEN full-length enriched, adult retina"
/tissue_type="retina"
/dev_stage="adult"

```

FEATURES

source

```

/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5'GAGAGAGAGATTCGAGTTAAATTAATTCCTCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I. -Retina RNA was provided by Stefano Gustincich,
Department of Neurobiology, Harvard Medical School, 220
Longwood Ave., Boston, MA02115, USA, whose assistance we
gratefully acknowledge."
BASE COUNT 159 a 168 c 185 g 150 t
ORIGIN
Query Match 8.5%; Score 242; DB 10; Length 662;
Best Local Similarity 60.5%; Pred. No. 11e-41;
Matches 398; Conservative 0; Mismatches 260; Indels 0; Gaps 0;
QY 778 AACCCCATCAACATCGTTGGTCAAGGTGCTGCTTCTTAGAGATCGTACTCGGCCCC 837
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5 AACCCATCCGCTGGCGGTGGTGAAGTGGTGTGCTGACGGACAAGGACAGAGTCTG 64
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 838 AAGTCAACGGCATCGGCCCTGACCTGGCAACTCTGTGCTGGGAGAGAAGCTG 897
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 65 GAGTGTAGCAAGAATGCGGCCGACGACCCCTCAAGAACTTTTGCAAAATGGCAGCACCAACAT 124
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 898 AACAAAGTGTAGTGACAAAGCACCCCGAGTACTGTGGGACACTGCCATCTCTTCCACGAGCAG 957
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 125 AACCACTAGGGATGATCAGGAAGACACTACGATCAGCCATCTGTTTCCACCCAGAG 184
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 958 GACCTGTGTGGGCCACCACTGTGACACCCCTGGGCATGGCTGATGGGTACCATGTGT 1017
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 185 GATTATGTGGGCATCATTTATGTGACACCCCTGGGAATGGCAGACGTTGGGACCATATGT 244
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1018 GACCCCAAGAGAAGCTGCTGTCTGTCATTGAGGACGATGGCTTCCATCAGCTTCCACCCT 1077
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 245 TCTCCGAGCGCAGCTGTGCAAGTATTGACATGATGGCCCTCCATGACGCTTTCATCTGTG 304
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1078 GCCCAGAGCTGGGCCACGTTTCAACAGCCCATGACAAATGTGAAGTCTGTGAGGAG 1137
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 305 GCTCATGAAATGGGCATCTACTTGGCCTTTCTCATGACGATTCCAAATCTCTGTAAGAG 364
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1138 GTGTTTGGGAAGTCCGAGCAACCAACATGATGTCCCGACCCCTCATCCAGATCGACGT 1197
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 365 AACTTCGGTACTACAGAAGACAAAGCGTTAATCTCTTCAATCTTACCAGCATCGATGCA 424
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1198 GCCAACCCCTGGTCAAGCTGCTGCTCATCATCAGCACTTCTCTGGACACGCGGCAC 1257
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 425 TCCAAGCCCTGGTCCAAATGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 484
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1258 GGTGACTGCTCTCTGGACCAACCCAGAGCCCATCTCCCTGCCCGAGGATCTGCCGGGC 1317
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 485 GGTAAATTTGTTGTAGACCTTACCACGGAAGACAGATTTTGGGTCCCGAGGAATCCCAAG 544
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1318 GCCAGCTACACCTGAGCCAGCAGTGGAGCTGGCTTTGGCGTGGCTCCCAAGCCCTGT 1377
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 545 CAGACCTACGATGGCCCGCAGCAGTGCACCTTACATTTGGGCTGAGTACTCGGTGTGC 604
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1378 CTTTACATGCACTACTGTCACCAAGCTGTGTGTCACCGGAAGGCCCAAGGACAGATGG 1435
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 605 CTTGGCATGATCTGTGCGCGCTGTGTGTGCTGTGTGTCGCCCGCAGGCAATTG 662
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```
RESULT 14
BQ930852      951 bp  mRNA  linear  EST 20-AUG-2002
LOCUS
DEFINITION   BQ930852.1 NCICGAP.Co24 Mus musculus cDNA clone
IMAGE:6395437 5', mRNA sequence.
ACCESSION   BQ930852
VERSION     BQ930852.1 GI:22345883
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 951)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13890 row: b column: 14
High quality sequence stop: 644.
FEATURES             source
    source
    1..951
        /organism="Mus musculus"
        /strain="FVB/N"
        /db_xref="taxon:10090"
        /clone_lib="IMAGE:6395437"
        /lab_host="NCI_CGAP-Co24"
        /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
        Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
        Average insert size 1.6 kb. Constructed by Life
        Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT          189 a 143 c 180 g 135 t
ORIGIN
Query Match      8.5%; Score 241.8; DB 14; Length 951;
Best Local Similarity 62.2%; Pred.No. 1.4e-41;
Matches 397; Conservative 0; Mismatches 238; Indels 3; Gaps 1;
Qy 1549 GGTTCCTGGGCAATGGGATCCCTATGGCCCTGCTCGCGACATCTGTGGGGGGCGTG 1608
Db 1 GGAAGCTGGGGACCATGGGACCGTGGGGAGACTGCTCAAGAACCTGTGTGGTGGAGTT 60
Qy 1609 CAGCTGGCCAGGAGCGAGTCGACCAACCCACCCTGCCAACGGGGCAAGTACTGCGAG 1668
Db 61 CAATACACAATGAGAGATGTGCAACACCCAGTCCCAAGAACGGGAGGAAGTACTGTGAA 120
Qy 1669 GGAGTGGGGTGAATACCGGATCTCGAACCTGGAGCCCTGCCACAGCTCAGCCTCCGGA 1728
Db 121 GGCAACGAGTCCGCTACAGTCTCTGTAACTCGAGGACTGTCCAGAC---AATAACGGA 177
Qy 1729 AAGAGCTTCGGGGAGGAGCGATGTGAGCTTTCAACGGCTTACACACAGCAACCAACCGG 1788
Db 178 AAAAGCTTCAGAGAGGAGCAGTGGCGGCGGACCAATGAGTTTCCAAAGCTTCTTTGGG 237
Qy 1789 CTCACCTCGCGCGTGGCATGGTGGCCCAAGTACTCCGCGGTGCTCCCGGGGCAAGTGC 1848
Db 238 AATGAGCCCACTGTAGAGTGAGACACCAACCAAGTACCCGCGCTCTCGCAAGGACAGTGC 297
Qy 1849 AAGCTCATCTCCGAGGCAATGGCACTGGCTACTTCTATGTGCTGGCAACCAAGGTGTTG 1908
Db 298 AAGCTCATCTGGAAGCAAAAGGCAATGGCTACTTTTTCGTTCTTACAGCCCAAGTTGTA 357
Qy 1909 GACGGCAGCTGTGCTTCCTGACTCCACCTCCGCTGTGTCTTCCAAAGCAAGTGCATCAAG 1968
Db 358 GATGSCACTCCCTGTAGTCCAGACTTACCTCTGCTGTGTGCAAGGCGAGTGTGTGAAA 417
```

```
Qy 1969 GCTGCTGTGATGGGAACCTGGGCTCCCAAGAGAGATTCGACAAAGTGTGGGTGTGGG 2028
Db 418 GCTGCTGTGATCGCATATAGACTCCAAAAGAGTTTGATAAGTGTGGGTGTGGGA 477
Qy 2029 GGAGACAATAAGAGTGCAGAAGGTGACTGGACTTTCACCAAGCCCATCATCGGTAC 2088
Db 478 GGAACGGTTCACATGCAAGAGATGTGAGGATAGTACATAGTACAAAGACCTGNGTAT 537
Qy 2089 AATTCGTGGTGGCCATCCCGCAGCGCCCTCAAGCATCGACATCCGCCAGCGGGTTAC 2148
Db 538 CATGACATGTGCACAAATTCCTGCTGGAGCCCAACATGAAGTGAACATCGGAATCAA 597
Qy 2149 AAAGGGCTGATCGGGGATGACAACTACCTGGCTCTGAA 2186
Db 598 AGGGGGTCCAGAAACAATGGCAGCTTCTGGCTATTAA 635

RESULT 15
BQ033017      945 bp  mRNA  linear  EST 20-OCT-2000
LOCUS
DEFINITION   BQ033017 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3859104 5',
mRNA sequence.
ACCESSION   BQ033017
VERSION     BQ033017.1 GI:10740729
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 945)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9592 row: b column: 01
High quality sequence stop: 642.
FEATURES             source
    source
    1..945
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:3859104"
        /clone_lib="NIH_MGC_66"
        /tissue_type="adenocarcinoma"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: ovary; Vector: pCMV-SPORT6; Site:1: NotI;
        Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
        Average insert size 1.8 kb. Library constructed by Life
        Technologies."
BASE COUNT          220 a 246 c 326 g 153 t
ORIGIN
Query Match      8.1%; Score 231.4; DB 12; Length 945;
Best Local Similarity 61.8%; Pred.No. 2.3e-39;
Matches 422; Conservative 0; Mismatches 251; Indels 10; Gaps 3;
Qy 1534 AACACAGGTTGATGTTCTTGGCCAAATGGATGCCCTATGTCCTGCTCGGCGACA 1593
Db 16 AGCCCCGTGTAGATGAGGCTGGGACCGCTGGGGACCTCGGGAGAAATGTTCTCGGACC 75
Qy 1594 TGTGGTGGGGCGTGCAGCTGGCCAGGAGGAGTGCACCAACCCCAACCCCTGCCAACGGG 1653
Db 76 TGTGGAGGAGGAGTACAGTTTTCACACCGCTGAGTGAAGGACCCCGAGCCTCAGATGA 135
Qy 1654 GGCAAGTACTGCGAGGAGTGGGGTGAATACCGGATCCTGCNAACCTGGAGCCCTGCCCC 1713
```

```

Db 136 GGAAGATACTGCTGGTGGTGGGAGAGCAAGTACCACTCATGCCACACGGAGGAATGCCOC 195
QY 1714 AGCTCAGCCTCCGAGAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTCACCGGCTACAC 1773
Db 196 -----CCTGACGGGAAAGCTTCAGGGAGCAGCAGTGTGAGAAGTATATGCCCTACAA 249
QY 1774 CACAGCACCAACCGGCTCACTCTGCGCGTGGCATGGGTGCCAAGTACTCCGGCGGTGTCT 1833
Db 250 TACACTGACATGGACGGGAATCT---CCTGCAGTGGGTCCCAAGTATGCTGGGTGTCC 306
QY 1834 CCGCGGCAAGTGCAGCTCATCTCCGAGCCAAATGGCACTGGCTACTTCTATGTGTG 1893
Db 307 CCGCGGAGCCGCTCAAGTGTGTCTCCGAGCCCGGGGAGGAGCGAGTCAAGGTGTC 366
QY 1894 GCACCCAAAGTGTGGACGCGCTGTCTCTCTGACTCCACCTCCGCTCTGTGTCCAA 1953
Db 367 GAGGCCAAAGTGTGATGATGCACCTGTGTGGGCCAGAAACACTGGCCATCTGTGTCCGT 426
QY 1954 GGCAAGTGCATCAAGGCTGGCTGTGATGGGAACCTGGGCTCCAAAGAGATTCGACAA 2013
Db 427 GGCCAGTGTGTCAAGGCCGGCTGTGACCATGTGGTGGACTCGCTCGGAAGCTGGACAAA 486
QY 2014 TGTGGGTGTGTGGGGGAGACAATAAGAGCTGCAAGAAGTGAAGTCTTTCACCAAG 2073
Db 487 TGGGGGTGTGTGGGGCAAGGCAACTCTCGAGGAGTCTCCGGGTCCCTCACCCCC 546
QY 2074 CCCATGCATGCTACAATTTCTGGTGGGCCATCCCGCAGCGCCCTCAAGCATCGACATC 2133
Db 547 ACCAATTATGCTACAATGACATTTGCACCATCCAGCTGGTGGCCACTAATATTGACGTG 606
QY 2134 CGCCAGCGCGTTCACAAAGGCTGATCGGGATGACAACTACCTGCTCTGAAGAACAGC 2193
Db 607 AAGCAGCG-GAGCACCCTGGTGTGCAGAACGATGGGACTACCTGGGGCTGAAGACGCT 665
QY 2194 CAAGGCAAGTACCTGCTCAACGG 2216
Db 666 GATGGCAGTACCTGCTCAAGGG 688

```

Search completed: May 1, 2003, 22:50:30
Job time : 3933 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 00:40:42 ; Search time 7405 Seconds
(without alignments)
11212.735 Million cell updates/sec

Title: us-10-009-332-2

Perfect score: 2853

Sequence: 1 atcttttctgggcatact.....gcgtctgagccgtgtgta.2853

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2746	96.2	2853	9	HSA315733	AJ315733 Homo sapi
2	2089	73.2	2930	6	AX342635	AX342635 Sequence
3	2089	73.2	2937	6	AX319860	AX319860 Sequence
4	902	31.6	157963	9	AP002986	AP002986 Homo sapi
5	902	31.6	172905	2	AC025130	AC025130 Homo sapi
6	803	28.1	170682	2	AC023429	AC023429 Homo sapi
7	775	27.2	28000	9	AP003459	AP003459 Homo sapi
8	358	12.5	172905	2	AC025130	AC025130 Homo sapi
9	288	10.1	170682	2	AC023429	AC023429 Homo sapi
10	74	2.6	182556	2	AC101990	AC101990 Mus muscu
11	61	2.1	178764	2	AC126507	AC126507 Rattus no
12	44	1.5	178764	2	AC126507	AC126507 Rattus no
13	36	1.3	182656	2	AC101990	AC101990 Mus muscu
14	35	1.2	3927	10	BC009667	BC009667 Mus muscu
15	23	0.8	86229	2	AC099229	AC099229 Rattus no
16	23	0.8	176054	2	AC022377	AC022377 Homo sapi
17	23	0.8	178548	9	AC018831	AC018831 Homo sapi
18	23	0.8	198597	9	AC026218	AC026218 Homo sapi
19	23	0.8	202844	9	AC090956	AC090956 Homo sapi
20	23	0.8	229402	9	AC069271	AC069271 Homo sapi
21	22	0.8	132131	9	AL139159	AL139159 Human DNA
22	22	0.8	172913	2	AC124158	AC124158 Macaca mu
23	21	0.7	1333	5	AF070952	AF070952 Xenopus l
24	21	0.7	1437	5	AF070953	AF070953 Xenopus l
25	21	0.7	1456	5	AF072582	AF072582 Xenopus l
26	21	0.7	1979	8	HVACXP113	X78877 Hordeum vul
27	21	0.7	2290	10	BC006797	BC006797 Mus muscu
28	21	0.7	3315	9	BC016649	BC016649 Homo sapi
29	21	0.7	7330	1	AB026907	AB026907 Ochrobact
30	21	0.7	10353	1	AE001884	AE001884 Deinococc
31	21	0.7	38393	9	HSN62C4	Z73963 Human DNA s
32	21	0.7	68882	2	AC125428	AC125428 Homo sapi
33	21	0.7	110000	2	AL583837_1	Continuation (2 of
34	21	0.7	117951	9	AL359821	AL359821 Human DNA
35	21	0.7	126315	8	AP004274	AP004274 Oryza sat
36	21	0.7	131121	8	AC092553	AC092553 Oryza sat
37	21	0.7	134609	2	AC021416	AC021416 Homo sapi
38	21	0.7	135245	9	AL136092	AL136092 Human DNA
39	21	0.7	137724	8	AC122147	AC122147 Oryza sat
40	21	0.7	147207	2	AC078890	AC078890 Oryza sat
41	21	0.7	152120	9	AL391294	AL391294 Human DNA
42	21	0.7	182041	2	AC023025	AC023025 Homo sapi
43	21	0.7	182916	9	AC112482	AC112482 Homo sapi
44	21	0.7	196378	2	AC117504	AC117504 Homo sapi
45	21	0.7	208695	10	AL671478	AL671478 Mouse DNA

ALIGNMENTS

RESULT 1
HSA315733
Homo sapiens mRNA for metalloproteinase disintegrin 15 (ADAMTS15 gene).
2853 bp mRNA linear PRI 01-MAR-2002
HSA315733
Homo sapiens mRNA for metalloproteinase disintegrin 15 (ADAMTS15 gene).
AJ315733.1 GI:19171175
ADAMTS15 gene; disintegrin; metalloproteinase; thrombospondin.
human.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
Cal,S., Obaya,A.J., Llamazares,M., Garabaya,C., Quesada,V. and Lopez-Otin,C.

TITLE Cloning, expression analysis, and structural characterization of seven novel human ADAMTS, a family of metalloproteinases with disintegrin and thrombospondin-1 domains
JOURNAL Gene 283 (1-2), 49-62 (2002)
MEDLINE 21856482
PUBMED 11867212
REFERENCE 2 (bases 1 to 2853)
AUTHORS Cal.S.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2001) Cal S., Biochemistry and Molecular Biology, University of Oviedo, Campus del Cristo, Asturias. 33006, SPAIN
FEATURES
source Location/Qualifiers
1. .2853
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1. .2853
/gene="ADAMTS15"
CDS 1. .2853
/gene="ADAMTS15"
/codon_start=1
/product="metalloprotease disintegrin 15 with thrombospondin domains"
/protein_id="CAC86014.1"
/db_xref="GI:19171176"
/translation="MLLLGLLLAFAGRTAGGSEPEREVVVPIRLDPDINGRRYWRG PEDSGDGLIFQITAFQEDFYHLHTDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCF YSGDVNAEDSFAVSLGGLRGAFYRGAEXVISPLNASAPAOINSQGAHLQRR YGPGSPGSDPTSCVSGWNPAILRLADPKPRAGFGESRRSRGRKRFVSIPIR YVTLVAADDESVMKFGADLEHVLTLATAARLXHPHSILNPINIVVKKVLLLRDR SGPKVTGNALTLRNCAMOKLKNVSKDHPYWDPAIFLTRQDLGGATCTCLGMAD VTWCDFPKSCVIEDGLPFAFTTAHELGHVFNMPHDNVKCEEVFGKLRANHMSP TLIQIDRANPWSASAIITFDLSGDLCLLDQPSKPSLPELDLPGLPSTUSQCEL AFVGSKPCPYMOYCTKLWCTKRAKQMYCOTRHPFADGTSCGEGKLCGLKACVERH NLNKRHVDGSKWAKDYPGCSRTCCGGVOLARQCTNPANGKRYCEGVKYRSCN LEPCSSAGSKSPREBQCAFNGYNHSTNRLTLAVAWPKYSGVSPDKCKLICRANG TGFFYLAPKVDGTLCSPDSTSVQGVKCIKAGCDNGLSKKRPDKCVCGDKNKSC KVTGGLTFPMHGYNVVAIPAGASSIDIRORGYKGLIGDDNYLALKNSQKYLNGH FVYSAVERDLVWKSGLLRYSGTGATVESLQASRPILPELITVEVLSVGMKTPRVRYSF YLPKPREDKSHKDPGRPSVLNLSVLSNOVEQDPDRPFARWVWAGSWGPCSASCG SGLQKRVDCPCRGSGAGRTVPACDAHRPVTQACGEPCTWELSAWSPCKSGRGFG RSLKVCVGHGRLLRDQNLHRLKPOELDFCVLRPC"
BASE COUNT 522 a 921 c 903 g 507 t
ORIGIN
Query Match 96.2%; Score 2746; DB 9; Length 2853;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2796; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 57 TGAGCCAGAGCGGAGGTAGTCGTTCCATCCGACTCGACCCGACATTAACGGCCGCG 116
Db 57 TGAGCCAGAGCGGAGGTAGTCGTTCCATCCGACTCGACCCGACATTAACGGCCGCG 116
Qy 117 CTACTACTGCGGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTTCAGATCACAGC 176
Db 117 CTACTACTGCGGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTTCAGATCACAGC 176
Qy 177 ATTTACAGAGAGACTTTTACCTACACTGACCGCGGATGCTCAGTTCTTGCGTCCGCGCTT 236
Db 177 ATTTACAGAGAGACTTTTACCTACACTGACCGCGGATGCTCAGTTCTTGCGTCCGCGCTT 236
Qy 237 CTCACACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTTGG 296
Db 237 CTCACACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTTGG 296
Qy 297 ACGTGTCTTCTATCTGGGACGTGAACCGCGAGCCGAGACTCGTTCGCTGCTGTGAGCCT 356
Db 297 ACGTGTCTTCTATCTGGGACGTGAACCGCGAGCCGAGACTCGTTCGCTGCTGTGAGCCT 356
Qy 357 GTGGGGGGGCTCCGCGAGCCTTTGGCTACCGAGGCGCGGAGTATGTCTATTAGCCCGCT 416
Db 357 GTGGGGGGGCTCCGCGAGCCTTTGGCTACCGAGGCGCGGAGTATGTCTATTAGCCCGCT 416
Qy 417 GCCCAATGCTAGCGCGCGCGGCGCACACAGCAGCGGCGCACACCTTCTTCCAGCG 476

Db 417 GCCCAATGCTAGCGCGCGCGGCGCACACAGCAGCGGCGCACACCTTCTTCCAGCG 476
Qy 477 CCGGGGTGTTCCGGCGGGGCGCTTCCGGAGAGCCCACTCTCTGCTGCGGGGTGGGCTCGGG 536
Db 477 CCGGGGTGTTCCGGCGGGGCGCTTCCGGAGAGCCCACTCTCTGCTGCGGGGTGGGCTCGGG 536
Qy 537 CTGGAACCCCGCCATCTTACGGGCGCTTGAGCCCTTACAGCCGGGCGGGGTGGGCTTCCG 596
Db 537 CTGGAACCCCGCCATCTTACGGGCGCTTGAGCCCTTACAGCCGGGCGGGGTGGGCTTCCG 596
Qy 597 GGAGAGTCTGAGCCGGCGAGTCTGGGCGGCGCAAGGCTTTCGTGTCTATCCCGCGGTA 656
Db 597 GGAGAGTCTGAGCCGGCGAGTCTGGGCGGCGCAAGGCTTTCGTGTCTATCCCGCGGTA 656
Qy 657 CGTGAGAGCGTGTGTGTCGCGGACGAGTCAATGTGTCGAAGTTCCACGGCGGGGAGCTTGA 716
Db 657 CGTGAGAGCGTGTGTGTCGCGGACGAGTCAATGTGTCGAAGTTCCACGGCGGGGAGCTTGA 716
Qy 717 ACATATATCTGTCGCTGCTGGCAACGGCGGCGGAGCTTACCGCCATCCAGCATCTCT 776
Db 717 ACATATATCTGTCGCTGCTGGCAACGGCGGCGGAGCTTACCGCCATCCAGCATCTCT 776
Qy 777 CAACCCCATCAACATCGTGTGTGTCGAAGTGTCTTCTTAGAGATCTGACTCCGGGCC 836
Db 777 CAACCCCATCAACATCGTGTGTGTCGAAGTGTCTTCTTAGAGATCTGACTCCGGGCC 836
Qy 837 CAAGGTACCCGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGGCAGAAAGCT 896
Db 837 CAAGGTACCCGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGGCAGAAAGCT 896
Qy 897 GAACAAAGTGTGACAGCAACCGCCGAGTACTGGGACACTGCCATCTCTTCCACAGCA 956
Db 897 GAACAAAGTGTGACAGCAACCGCCGAGTACTGGGACACTGCCATCTCTTCCACAGCA 956
Qy 957 GGACCTGTGTGAGCACCACCTGTGACACCTGGGCGATGGCTGATGTGGGTACCATGTG 1016
Db 957 GGACCTGTGTGAGCACCACCTGTGACACCTGGGCGATGGCTGATGTGGGTACCATGTG 1016
Qy 1017 TGACCCCAAGAGAAGTGTCTGTCATTGAGGACGATGGGCTTCCATCAGCCTTCCACCAC 1076
Db 1017 TGACCCCAAGAGAAGTGTCTGTCATTGAGGACGATGGGCTTCCATCAGCCTTCCACCAC 1076
Qy 1077 TGCCCAAGAGTGGGCCAGCTGTTCAAATGCCCCATGACAATGTGAAGTCTGTGAGGA 1136
Db 1077 TGCCCAAGAGTGGGCCAGCTGTTCAAATGCCCCATGACAATGTGAAGTCTGTGAGGA 1136
Qy 1137 GGTGTTTGGGAAGCTCCGAGGCCAACACATGATGTCCCCGACCCCTCATCCAGATCGACG 1196
Db 1137 GGTGTTTGGGAAGCTCCGAGGCCAACACATGATGTCCCCGACCCCTCATCCAGATCGACG 1196
Qy 1197 TGCCAAACCCCTGGTCAGCTGCTGCCATCATACCCGACTTCTTGGACAGCGGCA 1256
Db 1197 TGCCAAACCCCTGGTCAGCTGCTGCCATCATACCCGACTTCTTGGACAGCGGCA 1256
Qy 1257 CGGTGACTCTCTCGGCAACCCAGCAAGCCCATCTCCCTGCGCCGAGGATCTCCCGGG 1316
Db 1257 CGGTGACTCTCTCGGCAACCCAGCAAGCCCATCTCCCTGCGCCGAGGATCTCCCGGG 1316
Qy 1317 CGCAGCTACACCCCTGAGCCAGCAGTGGCGTGGCTTTTGGCGTGGGCTCCAAAGCCCTG 1376
Db 1317 CGCAGCTACACCCCTGAGCCAGCAGTGGCGTGGCTTTTGGCGTGGGCTCCAAAGCCCTG 1376
Qy 1377 TCTTTACATGACAGTACTGACCAAGCTGTGGTGCACCCGGAAGGCCAAGGGACAGATGT 1436
Db 1377 TCTTTACATGACAGTACTGACCAAGCTGTGGTGCACCCGGAAGGCCAAGGGACAGATGT 1436
Qy 1437 GTGCCAGACCCGCACTTCCCTCTGGGCCCATGGCACAGCTGTGGCGAGGGCAAGCTCTG 1496
Db 1437 GTGCCAGACCCGCACTTCCCTCTGGGCCCATGGCACAGCTGTGGCGAGGGCAAGCTCTG 1496
Qy 1497 CCTCAAAGGGGCGTGGTGGAGAGACAACTCTCAAAAGCACAGGCTGGATGGTTCCTG 1556
Db 1497 CCTCAAAGGGGCGTGGTGGAGAGACAACTCTCAAAAGCACAGGCTGGATGGTTCCTG 1556

Db 1497 CCTCAAAGGGGCTGCTGGAGAGACACAACCTCAACAAGCACAGGCTGGATGTTCTCTG 1556
QY 1557 GSCCAATGGAGTCCCTATGATGCCCTGCTCGGCACATGTTGGTGGGCGTGCAGCTGGC 1616
Db 1557 GSCCAATGGAGTCCCTATGATGCCCTGCTCGGCACATGTTGGTGGGCGTGCAGCTGGC 1616
QY 1617 CAGGAGGAGTGCACAAACCCACCCCTGCCAAGGAGGAGTGCAGGAGGAGTGCAG 1676
Db 1617 CAGGAGGAGTGCACAAACCCACCCCTGCCAAGGAGGAGTGCAGGAGGAGTGCAG 1676
QY 1677 GGTGAATACCGATGCTGCAACCTGGAGCCCTGCCAGCTAGCTCGCGAAAGAGCTT 1736
Db 1677 GGTGAATACCGATGCTGCAACCTGGAGCCCTGCCAGCTAGCTCGCGAAAGAGCTT 1736
QY 1737 CGGAGGAGAGTGTGAGGCTTCAACGCTACAACACAGCACCACCGCTCACTCT 1796
Db 1737 CGGAGGAGAGTGTGAGGCTTCAACGCTACAACACAGCACCACCGCTCACTCT 1796
QY 1797 CGCCTGGCATGGTGCACCAAGTACTCCGGGCTGTCTCCCGGACCAAGTGCACAGCTCAT 1856
Db 1797 CGCCTGGCATGGTGCACCAAGTACTCCGGGCTGTCTCCCGGACCAAGTGCACAGCTCAT 1856
QY 1857 CTGCCAGCAATGCGATGCTTCTATGCTGCGACCCCAAGTGTGGAGCGGAC 1916
Db 1857 CTGCCAGCAATGCGATGCTTCTATGCTGCGACCCCAAGTGTGGAGCGGAC 1916
QY 1917 GCTGTGCTCTCTGACTCCACTCGCTGTGTCACAGGCAAGTGCATCAAGGCTGCTG 1976
Db 1917 GCTGTGCTCTCTGACTCCACTCGCTGTGTCACAGGCAAGTGCATCAAGGCTGCTG 1976
QY 1977 TGATGGGAACCTGGGCTCCAAAGAGATTCGACAAGTGTGGGCTGTGTGGGGGAGACAA 2036
Db 1977 TGATGGGAACCTGGGCTCCAAAGAGATTCGACAAGTGTGGGCTGTGTGGGGGAGACAA 2036
QY 2037 TAAGAGTGCAGAGAGTGTGACTCTTCAACAGCCCATGATGCTCAATTTGCT 2096
Db 2037 TAAGAGTGCAGAGAGTGTGACTCTTCAACAGCCCATGATGCTCAATTTGCT 2096
QY 2097 GTGTGCCATCCCGCAGCGCTCAAGCATCGACATCCGACGCGCTTACAAGGCT 2156
Db 2097 GTGTGCCATCCCGCAGCGCTCAAGCATCGACATCCGACGCGCTTACAAGGCT 2156
QY 2157 GATCGGGATGACAACTACCTGGCTCTGAAGAACACCAAGCAAGTACCTGCTCAAGCG 2216
Db 2157 GATCGGGATGACAACTACCTGGCTCTGAAGAACACCAAGCAAGTACCTGCTCAAGCG 2216
QY 2217 GCATTTGCTGTGCTGGGCTGGAGCGGACCTGTGTGTGAAGGCACTGCTGCGGTA 2276
Db 2217 GCATTTGCTGTGCTGGGCTGGAGCGGACCTGTGTGTGAAGGCACTGCTGCGGTA 2276
QY 2277 CAGCGCACGGGCACAGCGGTGGAGAGCTGCAGGCTTCCCGGCCATCCTGGAGCGCT 2336
Db 2277 CAGCGCACGGGCACAGCGGTGGAGAGCTGCAGGCTTCCCGGCCATCCTGGAGCGCT 2336
QY 2337 GACGCTGGAGTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCCTTCTA 2396
Db 2337 GACGCTGGAGTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCCTTCTA 2396
QY 2397 TCTGCCAAGAGCTCGGGAGGACAAGTCTCTCATCCCAAGGACCCCGGGGACCTC 2456
Db 2397 TCTGCCAAGAGCTCGGGAGGACAAGTCTCTCATCCCAAGGACCCCGGGGACCTC 2456
QY 2457 TGTCTTGCAACAGCGCTCTCAGGCTCTCCAAACAGGTGGAGAGCGGAGCAGAGGCC 2516
Db 2457 TGTCTTGCAACAGCGCTCTCAGGCTCTCCAAACAGGTGGAGAGCGGAGCAGAGGCC 2516
QY 2517 CCTGCAAGCTGGGTGGTGGAGCTGGGGGCGGCTGCTCCGCGAGCTCGCGAGTGGGCT 2576
Db 2517 CCTGCAAGCTGGGTGGTGGAGCTGGGGGCGGCTGCTCCGCGAGCTCGCGAGTGGGCT 2576
QY 2577 GCAGAGCGGGGCTGGAGTCCGGGGCTCCCGGGGAGCGCAGCGGCTCCCTGCTGTGA 2636
Db 2577 GCAGAGCGGGGCTGGAGTCCGGGGCTCCCGGGGAGCGCAGCGGCTCCCTGCTGTGA 2636

QY 2637 TGCAGCCCATGGCCCGTGGAGACACAAGCCTGCGGGAGAGCCCTGCCCACTGJGAGCT 2696
Db 2637 TGCAGCCCATGGCCCGTGGAGACACAAGCCTGCGGGAGAGCCCTGCCCACTGJGAGCT 2696
QY 2697 CAGCGCTGTGTACACCTGTCTCAAGAGCTGCGGCCGGGATTTTCAGAGCGGTCTACTAA 2756
Db 2697 CAGCGCTGTGTACACCTGTCTCAAGAGCTGCGGCCGGGATTTTCAGAGCGGTCTACTAA 2756
QY 2757 GTGTGTGGGCCACGGAGCGGCTGCTGCCCGGACCAAGTGCACCTGCACCGCAAGCC 2816
Db 2757 GTGTGTGGGCCACGGAGCGGCTGCTGCCCGGACCAAGTGCACCTGCACCGCAAGCC 2816
QY 2817 CCAGGAGCTGACCTTCTGCTCTGAGGCGGTGCTGA 2853
Db 2817 CCAGGAGCTGACCTTCTGCTCTGAGGCGGTGCTGA 2853
RESULT 2
AX342635 2930 bp DNA Linear PAT 12-JAN-2002
LOCUS Sequence 32 from Patent WO0198468.
DEFINITION AX342635
ACCESSION AX342635
VERSION AX342635.1 GI:18152032
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Yue, H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, J.,
Tribouley, C.M., Deleane, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A.,
Hafalia, A., Khan, F.A., Wallis, N.K., Yao, M.G., Lu, D.A., Patterson, C.,
Tang, Y.T., Walsh, R.T., Azimzai, Y., Rammakumar, J., Xu, Y. and Reddy, R.
JOURNAL Patent: WO 0198468-A 32 27-DEC-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source 1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 7473089CB1"
BASE COUNT 529 a 958 c 931 g 512 t
ORIGIN
Query Match 73.2%; Score 2089; DB 6; Length 2930;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2379; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
QY 57 TGAGCCAGAGGGAGTACTGTTCCATCCGACTCGAGCCCGGACATTAACGGCGCGC 116
Db 131 TGAGCCAGAGGGAGTACTGTTCCATCCGACTCGAGCCCGGACATTAACGGCGCGC 190
QY 117 CTACTACTGGGGGTCGCCGAGGACTCCGGGATCAGGAGCTCATTTTCAGATACAGC 176
Db 191 CTACTACTGGGGGTCGCCGAGGACTCCGGGATCAGGAGCTCATTTTCAGATACAGC 250
QY 177 ATTTTCAGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCCTT 236
Db 251 ATTTTCAGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCCTT 310
QY 237 CTCACATGAGATCTGGGCTGCCCTCCAGGGGCTCACCGGGGCTCTTACAGCTGCG 296
Db 311 CTCACATGAGATCTGGGCTGCCCTCCAGGGGCTCACCGGGGCTCTTACAGCTGCG 370
QY 297 ACCTGCTTCTATTTCTGGGACCTGAACCGCGGAGCTCGTCTGCTCTGCTGAGCCT 356
Db 371 ACCTGCTTCTATTTCTGGGACCTGAACCGCGGAGCTCGTCTGCTCTGCTGAGCCT 430
QY 357 GTGCGGGGGCTCGCGAGCTTTGGCTACCGAGGCGCGAGTATGCTCTTACGCTCGCT 416
Db 431 GTGCGGGGGCTCGCGAGCTTTGGCTACCGAGGCGCGAGTATGCTCTTACGCTCGCT 490
QY 417 GCCCAATGCTAGCGCGCGCGCGCAGCAGCCAGCGCAGCCAGGCGCAGACCTTCTCCAGCG 476

	Db	491	GCCCAATGCTAGCGGGCGGGCGAGCGCAACGCCAGGCGCACACTTCTCACCAGC	550
	Qy	477	CCGGGTGTTCCGGCGGGCGCTTCCTCGGAGAACCCACCTCTCGCTCGGGGTGGCTCCGG	536
	Db	551	COGGGTGTTCCGGCGGGCGCTTCCTCGGAGAACCCACCTCTCGCTCGGGGTGGCTCCGG	610
	Qy	537	CTGGNACCCC GCCATCCTACGGGCCTTGAGCCCTTACAAGCCGCGCGGGGGTTCGG	596
	Db	611	CTGGNACCCC GCCATCCTACGGGCCTTGAGCCCTTACAAGCCGCGCGGGGGTTCGG	670
	Qy	597	GGAGAGTCGTAGCCGCGCGCAGGTCTGGCGCGCCAAAGCGTTTTCTATCCCGCGGTA	656
	Db	671	GGAGAGTCGTAGCCGCGCGCAGGTCTGGCGCGCCAAAGCGTTTTCTATCCCGCGGTA	730
	Qy	657	CGTGGAGACGCTGTTGTTCCGGACGAGTCAATGTFCAAAGTTCCAAGCGCGGACCTTGA	716
	Db	731	CGTGGAGACGCTGTTGTTCCGGACGAGTCAATGTTCAAAGTTCCAAGCGCGGACCTTGA	790
	Qy	717	ACATTATCTGCTGACGCTGCTGCAACGCGCGGGGACTCTACCGCCATCCACGATCCCT	775
	Db	791	ACATTATCTGCTGACGCTGCTGCAACGCGCGGGGACTCTACCGCCATCCACGATCCCT	850
	Qy	777	CACCCCATCAACATCTGTTGTTCAAGTGTGTTCTTAGAGATCTGACTCCGGGCC	836
	Db	851	CAACCCCATCAACATCTGTTGTTCAAGTGTGTTCTTAGAGATCTGACTCCGGGCC	910
	Qy	837	CAAGGTCACCGGAATCCGCCCTGACGCTGGCAACTTCTGTGCTGGCAGAGAAGACT	896
	Db	911	CAAGGTCACCGGAATCCGCCCTGACGCTGGCAACTTCTGTGCTGGCAGAGAAGACT	970
	Qy	897	GAACAAAGTGAGTGACAAAGCACCCCGAGTACTGGGACACTGCCATCTCTTCAACAGCA	956
	Db	971	GAACAAAGTGAGTGACAAAGCACCCCGAGTACTGGGACACTGCCATCTCTTCAACAGCA	1030
	Qy	957	GGACTGTGTGGAGCACACACTGTGACACCCCTGGGCATGGCTGATGGGTACCATGTG	1016
	Db	1031	GGACTGTGTGGAGCACACACTGTGACACCCCTGGGCATGGCTGATGGGTACCATGTG	1090
	Qy	1017	TGACCCCAAGAGAAGTGTCTGTCTAATCAGACACATGGGCTTCCATCAGCTTTCACAC	1076
	Db	1091	TGACCCCAAGAGAAGTGTCTGTCTAATCAGACACATGGGCTTCCATCAGCTTTCACAC	1150
	Qy	1077	TGCCCCAGAGCTGGGCCAGTGTTCACATGCCCATGACAAATGTGAAAGTCTGTGAGGA	1136
	Db	1151	TGCCCCAGAGCTGGGCCAGTGTTCACATGCCCATGACAAATGTGAAAGTCTGTGAGGA	1210
	Qy	1137	GGTGTGTTGGGAAGCTCCGAGCAACACATGATGTCCCGACCTCATCCAGATGACCG	1196
	Db	1211	GGTGTGTTGGGAAGCTCCGAGCAACACATGATGTCCCGACCTCATCCAGATGACCG	1270
	Qy	1197	TGCCAACCCCTGTGACAGCTGACGTGCCATCATCACCAGCTTCTTGACAGCGGCA	1256
	Db	1271	TGCCAACCCCTGTGACAGCTGACGTGCCATCATCACCAGCTTCTTGACAGCGGCA	1330
	Qy	1257	CGGTGACTGCTCCTTGACCAACCCAGACCCCATCTCCCTGGCCGAGGATGTCGCGG	1316
	Db	1331	CGGTGACTGCTCCTTGACCAACCCAGACCCCATCTCCCTGGCCGAGGATGTCGCGG	1390
	Qy	1317	CGCCAGCTACACCTTGAGCCAGCAGTGGCGTGGCTTTTGGCGTGGGCTCCAGAGCCCTG	1376
	Db	1391	CGCCAGCTACACCTTGAGCCAGCAGTGGCGTGGCTTTTGGCGTGGGCTCCAGAGCCCTG	1450
	Qy	1377	TCCTTAGATCGACTGTGACCAAGCTGTGTGTGACACCGGGAAGGCCAAGGACAGATGGT	1436
	Db	1451	TCCTTAGATCGACTGTGACCAAGCTGTGTGTGACACCGGGAAGGCCAAGGACAGATGGT	1510
	Qy	1437	GTGCCAGACCCGCACTTCCCTGGGCCGATGGCACACAGCTGTGGCGAGGCAAGCTCTG	1496
	Db	1511	GTGCCAGACCCGCACTTCCCTGGGCCGATGGCACACAGCTGTGGCGAGGCAAGCTCTG	1570
	Qy	1497	CCTCAAGGGGCTGCGTGGAGAGACAAACCTCAACAGCACAGGTTGATGGTTCTCTG	1556

Db	1571	CCTCAAAGGGGCCCTGCCTGGAGAGACACAACCTCAACAAAGACACAGGGTGGATGGTTCTCCTG	1630
Qy	1557	GGCCAAATGGGATCCCTATATGGCCCTGCTCGCACATGCTGCTGGGGCGTGCAGCTGGC	1616
Db	1631	GGCCAAATGGGATCCCTATATGGCCCTGCTCGGCACATGCTGCTGGGGCGTGCAGCTGGC	1650
Qy	1617	CAGGAGGAGTGCACAAACCCCAACCCCTGCGCAACGGGGGCAAGTACTCGGAGGGAGTGAG	1676
Db	1691	CAGGAGCAGTGCACCAACCCCAACCCCTGCCAACGGGGCAAGTACTCGGAGGGAGTGAG	1750
Qy	1677	GGTGAATACCGATCTCTGCAACCTGGAGCCCTGCCCAAGCTCAGCTCCCGAANAAGACTT	1736
Db	1751	GGTGAATACCGATCTCTGCAATCTGGAGCCCTGCCCAAGCTCAGCTCCCGAANAAGACTT	1810
Qy	1737	CCGGAGGAGCAGTGGAGGCTTTCAACGGCTACAACACAGACCAACACCGGCTCACTCT	1796
Db	1811	CCGGAGGAGCAGTGGAGGCTTTCAACGGCTACAACACAGACCAACACCGGCTCACTCT	1870
Qy	1797	CGCGTGCATGGGTGCCAAGTACTCCGGCGTGTCTCCCGGGACAAGTGCACAGCTCAT	1856
Db	1871	CGCGTGGCATGGGTGCCAAGTACTCCGGCGTGTCTCCCGGGACAAGTGCACAGCTCAT	1930
Qy	1857	CTCGCGACCAATGGCACTGGCTACTTCTATGTGTGGCACCCAA---GGTGGTGGACGG	1913
Db	1931	CTCGCGACCAATGGCACTGGCTACTTCTATGTGTGGCACCCAAAGTGTGTGGTGGACGG	1990
Qy	1914	CACGCTGTCTCTCGTACATCCACCTCCGCTGTGTCCAAGGCAAGTGCATCAAGGCTGG	1973
Db	1991	CACGCTGTCTCTCGTACATCCACCTCCGCTGTGTCCAAGGCAAGTGCATCAAGGCTGG	2050
Qy	1974	CTGTATGGGAACCTGGGCTCCAAGAAGAGATTCACAAAGTGTGGGGTGTGTGGGGGAGA	2033
Db	2051	CTGTATGGGAACCTGGGCTCCAAGAAGAGATTCACAAAGTGTGGGGTGTGTGGGGGAGA	2110
Qy	2034	CAATAAGAGCTGCAGNAGTGAATGGACTCTTCCAAAGCCCATGCATGCTACAATTT	2093
Db	2111	CAATAAGAGCTGCAGNAGTGAATGGACTCTTCCAAAGCCCATGCATGCTACAATTT	2170
Qy	2094	CGTGGTGCCATCCCGCAGGCGCCTCAAGCATCGCACATCCGCACGCGGTTCACAAAGG	2153
Db	2171	CGTGGTGCCATCCCGCAGGCGCCTCAAGCATCGCACATCCGCACGCGGTTCACAAAGG	2230
Qy	2154	GCTATCGGGATGCACACTACTGCTCTGAAGACAGCCAAAGCAGTACCTGCTCAA	2213
Db	2231	GCTATCGGGATGCACACTACTGCTCTGAAGACAGCCAAAGCAGTACCTGCTCAA	2290
Qy	2214	CGGCAATTTCTGTGTCTGTCGGCGGTGGAGCGGACCTGTGTGAAGGSCAGTCTGCTGCG	2273
Db	2291	CGGCAATTTCTGTGTCTGTCGGCGGTGGAGCGGACCTGTGTGAAGGSCAGTCTGCTGCG	2350
Qy	2274	GTACAGCGGCACGGGCACACGGTGGAGAGCTTGACGGCTTCCCGGCCCATCTCTGGAGCC	2333
Db	2351	GTACAGCGGCACGGGCACACGGTGGAGAGCTTGACGGCTTCCCGGCCCATCTCTGGAGCC	2410
Qy	2334	GCTGACCGTGGAGGTCTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCCTT	2393
Db	2411	GCTGACCGTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCCTT	2470
Qy	2394	CTATCTGCCCAAGAGCCTCGGAGAGCAAGTCTCTCTCATCCC	2436
Db	2471	CTATCTGCCCAAGAGCCTCGGAGAGCAAGTCTCTCTCATCCC	2513

RESULT 3	
AX319860	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

AX319860 2937 bp
Sequence 24 from Patent WO0183782.

ACCESSION AX319860
VERSION AX319860.1 GI:17901450
KEYWORDS
SOURCE human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia.

REFERENCE 1
AUTHORS Plozman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and Payne,V.
TITLE Novel proteases
JOURNAL Patent: WO 0183782-A 24 08-NOV-2001;
Sugen, Inc. (US)
FEATURES Location/Qualifiers
source 1..2937
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 531 a 958 c 936 g 512 t
ORIGIN

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match 73.2%; Score 2089; DB 6; Length 2937;

Best Local Similarity 99.8%; Pred No. 0;

Matches 2379; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY	57	TGAGCCAGAGCGGAGGTAGTCTCCCATCCGACTGGACCCGGACATTAAACGGCGCGG	116
DB	132	TGAGCCAGAGCGGAGGTAGTCTCCCATCCGACTGGACCCGGACATTAAACGGCGCGG	191
QY	117	CTACTACTGCGGGGTCGCGAGGACTCCGGGATCAGGACTCATTTTCAGATCACAGC	176
DB	192	CTACTACTGCGGGGTCGCGAGGACTCCGGGATCAGGACTCATTTTCAGATCACAGC	251
QY	177	ATTTCAGGAGGACTTTTACCTACACCTAGCGCGGATGCTCAGTTCTTGGCTCCCGCCTT	236
DB	252	ATTTCAGGAGGACTTTTACCTACACCTAGCGCGGATGCTCAGTTCTTGGCTCCCGCCTT	311
QY	237	CTCCACTGAGCATCTGGGGTCTCCCTCCAGGGGCTACCGGGGCTCTTCAGACCTGCG	296
DB	312	CTCCACTGAGCATCTGGGGTCTCCCTCCAGGGGCTACCGGGGCTCTTCAGACCTGCG	371
QY	297	ACGCTGCTTCTATTCTGGGAGGTGAACCGCGAGCGGACTGCTTCGCTGCTGAGCCT	356
DB	372	ACGCTGCTTCTATTCTGGGAGGTGAACCGCGAGCGGACTGCTTCGCTGCTGAGCCT	431
QY	357	GTGCGGGGGGCTCCGGGAGGCTTTGGCTACCGAGCGCGGAGTAGTATGTCATTAGCCGCT	416
DB	432	GTGCGGGGGGCTCCGGGAGGCTTTGGCTACCGAGCGCGGAGTAGTATGTCATTAGCCGCT	491
QY	417	GCCCAATGCTAGCGCGCGCGGCGCAGCGGCAACAGCCAGCGGACACCTTCTCCAGG	476
DB	492	GCCCAATGCTAGCGCGCGCGGCGCAGCGGCAACAGCCAGCGGACACCTTCTCCAGG	551
QY	477	CCGGGGTGTTCGGGGGGGCTTCCGGAGACCCACCTCTCGCTGCGGGGTGGCTCGGG	536
DB	552	CCGGGGTGTTCGGGGGGGCTTCCGGAGACCCACCTCTCGCTGCGGGGTGGCTCGGG	611
QY	537	CTGGAACCCCGCATCTTACGGGCGCTTGACCCCTTACAGCCGCGGGGGGGGCTTCGG	596
DB	612	CTGGAACCCCGCATCTTACGGGCGCTTGACCCCTTACAGCCGCGGGGGGGGCTTCGG	671
QY	597	GGAGAGTCTAGCGCGCGAGTCTTGGGGCGCCAGGCTTTCGTTCTATCCCGCGGTA	656
DB	672	GGAGAGTCTAGCGCGCGAGTCTTGGGGCGCCAGGCTTTCGTTCTATCCCGCGGTA	731
QY	657	CGTGGAGACGCTGGTGGTCCGAGAGTCAATGGTCCAGTTCCAGCGCGGAGCTTGG	716
DB	732	CGTGGAGACGCTGGTGGTCCGAGAGTCAATGGTCCAGTTCCAGCGCGGAGCTTGG	791
QY	717	ACATTATCTGCTGACGCTGCTGGCAACGCGCGGAGCTCTACCGCATCCAGATCCT	776
DB	792	ACATTATCTGCTGACGCTGCTGGCAACGCGCGGAGCTCTACCGCATCCAGATCCT	851
QY	777	CAACCCCATCAACATGTTGTGCTCAAGTGCTCTTCTAGAGATCTGACTCCGGGCG	836
DB	852	CAACCCCATCAACATGTTGTGCTCAAGTGCTCTTCTAGAGATCTGACTCCGGGCG	911
QY	837	CAAGGTACCGGCAATGCGGGCCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAAGCT	896
DB	912	CAAGGTACCGGCAATGCGGGCCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAAGCT	971

QY	897	GAACAAATGAGTGACAAAGCACCCCGAGTACTGGGACACTGCCATCTCTTTCACAGGCA	956
DB	972	GAACAAATGAGTGACAAAGCACCCCGAGTACTGGGACACTGCCATCTCTTTCACAGGCA	1031
QY	957	GGACCTGTGTGAGGACCAACCTGTGACACCTCGGATGCTGATGGGTACCATGTG	1016
DB	1032	GGACCTGTGTGAGGACCAACCTGTGACACCTCGGATGCTGATGGGTACCATGTG	1091
QY	1017	TGACCCCAAGAGAAAGCTGCTGTCTGATTGAGGACGATGGGCTTCATCAGCTTTCACAC	1076
DB	1092	TGACCCCAAGAGAAAGCTGCTGTCTGATTGAGGACGATGGGCTTCATCAGCTTTCACAC	1151
QY	1077	TGCCACACAGCTGGCCACGCTTTCACATGCCCCCATGACAAATGTAAAGTCTGTGAGGA	1136
DB	1152	TGCCACACAGCTGGCCACGCTTTCACATGCCCCCATGACAAATGTAAAGTCTGTGAGGA	1211
QY	1137	GGTGTGTTGGGAAGCTCCGAGCAACACCATGATGTCCCGACCCCTCATCAGATCGACCG	1196
DB	1212	GGTGTGTTGGGAAGCTCCGAGCAACACCATGATGTCCCGACCCCTCATCAGATCGACCG	1271
QY	1197	TGCCAAACCCCTGTGTCAGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATG	1256
DB	1272	TGCCAAACCCCTGTGTCAGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATG	1331
QY	1257	CGGTGACTGCTCTCTGGACCAACACCATGATGTCCCGACCCCTCATCAGATCGACCG	1316
DB	1332	CGGTGACTGCTCTCTGGACCAACACCATGATGTCCCGACCCCTCATCAGATCGACCG	1391
QY	1317	CGCCAGCTTACACCTGAGCCAGCAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG	1376
DB	1392	CGCCAGCTTACACCTGAGCCAGCAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG	1451
QY	1377	TCCTTACATGACGCTACTGTCACCAAGCTGTGTGTGACCGGAGGCCAAGGACAGATGCT	1436
DB	1452	TCCTTACATGACGCTACTGTCACCAAGCTGTGTGTGACCGGAGGCCAAGGACAGATGCT	1511
QY	1437	GTGCGAGACCCCGCCACTTCCCTGTGGCGGATGGCACCAGCTGTGGCGAGGGCAAGCTCTG	1496
DB	1512	GTGCGAGACCCCGCCACTTCCCTGTGGCGGATGGCACCAGCTGTGGCGAGGGCAAGCTCTG	1571
QY	1497	CCTCAAGGGGCTGTGGTGAGACACACCTCAACAGCAGGCTGATGATGATGATGATGATGATG	1556
DB	1572	CCTCAAGGGGCTGTGGTGAGACACACCTCAACAGCAGGCTGATGATGATGATGATGATGATG	1631
QY	1557	GGCCAAATGGGATCCCTATGGCCCTGCTCGGCACATGTGGTGGGGGCTGGCTGGCTGGC	1616
DB	1632	GGCCAAATGGGATCCCTATGGCCCTGCTCGGCACATGTGGTGGGGGCTGGCTGGCTGGC	1691
QY	1617	CAGGAGGAGTGCACCAACCCCTGTCACCGGGGCAAGTACTGCGAGGAGTGCAG	1676
DB	1692	CAGGAGGAGTGCACCAACCCCTGTCACCGGGGCAAGTACTGCGAGGAGTGCAG	1751
QY	1677	GGTGAATACCGATCCCTGCAACCTGAGCCCTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG	1736
DB	1752	GGTGAATACCGATCCCTGCAACCTGAGCCCTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG	1811
QY	1737	CCGGGAGGAGCAGTGTGAGGCTTTCACAGGCTACACACAGCAGCAACACCGGCTCACTCT	1796
DB	1812	CCGGGAGGAGCAGTGTGAGGCTTTCACAGGCTACACACAGCAGCAACACCGGCTCACTCT	1871
QY	1797	CGCCGTGGCATGGGTGCCCAAGTACTCCGGGCTGTCTCCCGGGGACAAAGTGCAGGCTCAT	1856
DB	1872	CGCCGTGGCATGGGTGCCCAAGTACTCCGGGCTGTCTCCCGGGGACAAAGTGCAGGCTCAT	1931
QY	1857	CTGCCGAGCCAAATGGGACTGGCTACTTCTATGCTGGGACCCCA --- GGTGTGAGCGG	1913
DB	1932	CTGCCGAGCCAAATGGGACTGGCTACTTCTATGCTGGGACCCCAAGGCTGGTGGTGGACCG	1991
QY	1914	CACGCTGTGCTCTCTGACTCCACCTCCGCTGTGTGTGTCACAGGCAAGTGCATCAAGCTGG	1973
DB	1992	CACGCTGTGCTCTCTGACTCCACCTCCGCTGTGTGTGTCACAGGCAAGTGCATCAAGCTGG	2051

QY 1974 CTGTGATGGGAACCTGGGCTCCAGAGAGATTCGACAAAGTGTGGGTGTGGGGAGA 2033
Db 2052 CTGTGATGGGAACCTGGGCTCCAGAGAGATTCGACAAAGTGTGGGTGTGGGGAGA 2111
QY 2034 CAATAAGAGCTGCAAGAGGTGACTGGACTCTTCACCAAGCCCATGCATGGCTACAAATTT 2093
Db 2112 CAATAAGAGCTGCAAGAGGTGACTGGACTCTTCACCAAGCCCATGCATGGCTACAAATTT 2171
QY 2094 CGTGGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGACGCGGTTCACAAAGG 2153
Db 2172 CGTGGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGACGCGGTTCACAAAGG 2231
QY 2154 GCTGATCGGGGATGACAACTACCTGGCTCTGGAAGACAGCCAAAGCAAGTACCTGCTCAA 2213
Db 2232 GCTGATCGGGGATGACAACTACCTGGCTCTGGAAGACAGCCAAAGCAAGTACCTGCTCAA 2291
QY 2214 CGGCAATTCGTGTGCGCGGTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGG 2273
Db 2292 CGGCAATTCGTGTGCGCGGTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGG 2351
QY 2274 GTACAGCGGCACGGGACAGCGGTGGAGAGCCTGCAAGGCTTCCCGGCCCATCTGGAGCC 2333
Db 2352 GTACAGCGGCACGGGACAGCGGTGGAGAGCCTGCAAGGCTTCCCGGCCCATCTGGAGCC 2411
QY 2334 GCTGACCGTGGAGGTCTCTCCGTGGGGAGATGACACCGCCCGGCTCCGCTACTCCTT 2393
Db 2412 GCTGACCGTGGAGGTCTCTCCGTGGGGAGATGACACCGCCCGGCTCCGCTACTCCTT 2471
QY 2394 CTATCTGCCCAAGAGCCTCGGAGGACAAAGTCTCTCATCCC 2436
Db 2472 CTATCTGCCCAAGAGCCTCGGAGGACAAAGTCTCTCATCCC 2514

RESULT 4
AP002986 157963 bp DNA linear PRI 08-MAR-2002
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:RP11-121M22,
DEFINITION complete sequence.

ACCESSION AP002986
VERSION AP002986.2 GI:19263031
KEYWORDS HTG.

SOURCE Homo sapiens DNA, clone:RP11-121M22.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 157963)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission
JOURNAL Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT On Mar 7, 2002 this sequence version replaced gi:11559301.
FEATURES Location/Qualifiers

source
1..157963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-121M22"

BASE COUNT 39742 a 36722 c 37247 g 44252 t
ORIGIN

Query Match 31.6%; Score 902; DB 9; Length 157963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 TGAGCCAGAGCGGAGGTAGTCTCCATCCGACTGGACCCGACATTAACGCCGCCG 116
Db 134171 TGAGCCAGAGCGGAGGTAGTCTCCATCCGACTGGACCCGACATTAACGCCGCCG 134230
QY 117 CTACTACTGGGGGTCCGAGGACTCCGGGATCAGGAGTCAATTTTCAGATCAGAC 176
Db 134231 CTACTACTGGGGGTCCGAGGACTCCGGGATCAGGAGTCAATTTTCAGATCAGAC 134290
QY 177 ATTTTCAGAGGACTTTTACCTACACCTGACGCCGAGTGTCTCAGTTCTTGGCTCCCGGCTT 236
Db 134291 ATTTTCAGAGGACTTTTACCTACACCTGACGCCGAGTGTCTCAGTTCTTGGCTCCCGGCTT 134350
QY 237 CTCCACTCAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGGCTTTCAGACCTGGC 296
Db 134351 CTCCACTCAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGGCTTTCAGACCTGGC 134410
QY 297 AGCTGTCTTCTATCTGGGGAGCTGAACGCCGAGCCGAGTCTGCTGCTGTGTGAGCCT 356
Db 134411 AGCTGTCTTCTATCTGGGGAGCTGAACGCCGAGCCGAGTCTGCTGCTGTGTGAGCCT 134470
QY 357 GTGCGGGGGCTCCGCGGAGCTTTGGCTACCGAGGCCGAGTATGTCAATFAGCCGCT 416
Db 134471 GTGCGGGGGCTCCGCGGAGCTTTGGCTACCGAGGCCGAGTATGTCAATFAGCCGCT 134530
QY 417 GCCCAATCTAGCGCGCGGCGGCGAGCGCAACAGCAGCGGCGCACACTTCTCCAGCG 476
Db 134531 GCCCAATCTAGCGCGCGGCGGCGAGCGCAACAGCAGCGGCGCACACTTCTCCAGCG 134590
QY 477 CCGGGGTGTTCGGCGGCGGCTTCCGGAGACCCACCTCTCGCTCGGGGTGGCTCGGG 536
Db 134591 CCGGGGTGTTCGGCGGCGGCTTCCGGAGACCCACCTCTCGCTCGGGGTGGCTCGGG 134650
QY 537 CTGGAACCCCGCATCTACGGGCGCTTACAGCGCGGCGGCGGCGGCTTCGG 596
Db 134651 CTGGAACCCCGCATCTACGGGCGCTTACAGCGCGGCGGCGGCGGCTTCGG 134710
QY 597 GGAGAGTCTAGCGCGGCGGCGGCTTGGGCGGCGGCGGCGGCTTCTCCCGCGGTA 656
Db 134711 GGAGAGTCTAGCGCGGCGGCGGCTTGGGCGGCGGCGGCGGCTTCTCCCGCGGTA 134770
QY 657 CGTGAGAGCGTGTGGTGGCGGAGTCAATGTGTAAGTTCCACGGCGCGGCGGCTTGA 716
Db 134771 CGTGAGAGCGTGTGGTGGCGGAGTCAATGTGTAAGTTCCACGGCGCGGCGGCTTGA 134830
QY 717 ACATTATCTGTACGCTGTGGCAACGGCGGCGGCGGCTTACCGCCCATCCAGCATCCT 776
Db 134831 ACATTATCTGTACGCTGTGGCAACGGCGGCGGCGGCTTACCGCCCATCCAGCATCCT 134890
QY 777 CAACCCCATCAACATCGTGTGGTCAAGTGTCTTCTTAGAGATCGTACTCGGGCC 836
Db 134891 CAACCCCATCAACATCGTGTGGTCAAGTGTCTTCTTAGAGATCGTACTCGGGCC 134950
QY 837 CAAGTCAACGGCAATGGCGGCTGACCTGCGCAACTTCTGTGCTGGCAGAGAAGCT 896
Db 134951 CAAGTCAACGGCAATGGCGGCTGACCTGCGCAACTTCTGTGCTGGCAGAGAAGCT 135010
QY 897 GAACAAAGTGTAGTACAGACCCCGAGTACTGGGACACTGCGATCTCTTTCACAGGCA 956
Db 135011 GAACAAAGTGTAGTACAGACCCCGAGTACTGGGACACTGCGATCTCTTTCACAGGCA 135070
QY 957 GG 958
Db 135071 GG 135072

RESULT 5
AC025130/c 172905 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 11 clone RP11-211H6 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
ACCESSION AC025130
VERSION AC025130.2 GI:7387384


```
misc_feature 23636..30261
/clone="assembly_fragment
vector_side:right"
30362..40941
/notes="assembly_fragment"
41042..54893
/notes="assembly_fragment"
54994..68888
/notes="assembly_fragment"
68989..83702
/notes="assembly_fragment"
83803..98393
/notes="assembly_fragment"
98494..117356
/notes="assembly_fragment"
117457..140589
/notes="assembly_fragment"
140690..172905
/notes="assembly_fragment"
BASE COUNT 44360 a 41333 c 39516 g 46194 t 1502 others

Query Match 31.6%; Score 902; DB 2; Length 172905;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 TGACCCAGAGCGGAGTAGTCGTTCCATCCGACTGGACCCGACATTAACGGCCGCG 116
Db 79527 TGACCCAGAGCGGAGTAGTCGTTCCATCCGACTGGACCCGACATTAACGGCCGCG 79468;

QY 117 CTACTACTGGGGGTCGCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAGC 176
Db 79467 CTACTACTGGGGGTCGCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAGC 79408;

QY 177 ATTTTCAGAGGACTTTTACCTACACTGACGCGGAGTCTCAGTCTTGCTCCGCCCTT 236
Db 79407 ATTTTCAGAGGACTTTTACCTACACTGACGCGGAGTCTCAGTCTTGCTCCGCCCTT 79348;

QY 237 CTCACACTGAGCATCTGGCGCTCCGCCCTCCAGGGGCTCACCGGGGCTCTTTACAGACCTCGG 236
Db 79347 CTCACACTGAGCATCTGGCGCTCCGCCCTCCAGGGGCTCACCGGGGCTCTTTACAGACCTCGG 79288;

QY 297 ACGTGTCTTATTTCTGGGACGTGAACGCGGAGCGGAGCTGCTGCTGCTGAGCCT 356
Db 79287 ACGTGTCTTATTTCTGGGACGTGAACGCGGAGCGGAGCTGCTGCTGCTGAGCCT 79228;

QY 357 GTGCGGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGCTATTAGCCGCT 416
Db 79227 GTGCGGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGCTATTAGCCGCT 79168;

QY 417 GCCCAATGCTAGCGCGCGCGGCGGACGCAACAGCCAGGGGCGACACCTTCTCCAGCG 476
Db 79167 GCCCAATGCTAGCGCGCGCGGCGGACGCAACAGCCAGGGGCGACACCTTCTCCAGCG 79108;

QY 477 CCGGGGTGTTCCGGCGGCTTCCGAGACCCACCTCTCGTGGCGGCTGCTCGGG 536
Db 79107 CCGGGGTGTTCCGGCGGCTTCCGAGAGACCCACCTCTCGTGGCGGCTGCTCGGG 79048;

QY 537 CTGGAACCCGCCCATCTACGGGCGCTTGACCCCTTACAAGCCGCGGCGGCGGCTTCGG 596
Db 79047 CTGGAACCCGCCCATCTACGGGCGCTTGACCCCTTACAAGCCGCGGCGGCGGCTTCGG 78988;

QY 597 GGAGAGTGTAGCGGCGGAGGCTGGGCGGCCCAAGCGTTTGTGTATTCGCCGGGTA 656
Db 78987 GGAGAGTGTAGCGGCGGAGGCTGGGCGGCCCAAGCGTTTGTGTATTCGCCGGGTA 78928;

QY 657 CGTGAGACGCTGTGTGCGGACGAGTCAATGCTCAAGTTCACCGCGGCGGACCTGGA 716
Db 78927 CGTGAGACGCTGTGTGCGGACGAGTCAATGCTCAAGTTCACCGCGGCGGACCTGGA 78868;

QY 717 ACATTATCTGCTGCTGTGGCAAGCGGCGGCGGACTCTACGCCATCCACGATCCT 776
|||||
```

```
Db 78867 ACATTATCTGCTGAGCGCTGTGGCAACGGCGGCGGACTCTACCGCCATCCACGATCCT 78808
QY 777 CAACCCCATCAACATCGTGTGGTCAAGGTGCTGCTTCTTAGAGATCGTGACTCCGGGCC 836
|||||
Db 78807 CAACCCCATCAACATCGTGTGGTCAAGGTGCTGCTTCTTAGAGATCGTGACTCCGGGCC 78748
|||||
QY 837 CAAGGTACCGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAGCT 896
|||||
Db 78747 CAAGGTACCGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAGCT 78688
|||||
QY 897 GAACAAAGTGAAGTGAACAGCACCCGAGTACTGGGACACTGCCATCTCTTCCACAGCA 956
|||||
Db 78687 GAACAAAGTGAAGTGAACAGCACCCGAGTACTGGGACACTGCCATCTCTTCCACAGCA 78628
|||||
QY 957 GG 958
||
Db 78627 GG 78626

RESULT 6
AC023429
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-121M22, WORKING DRAFT
AC023429
AC023429.19 GI:13569974
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 170682)
AUTHORS Abola,A.P., Bruno,D., Conn,L., Della Rosa,M., Faulkner,D.,
Federici,N., Glukhov,S., Hansen,N., Herman,Z.S., Hymen,R.,
Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
Southwick,A.M., Webb,C., Wilhelm,J., Yu,S. and Davis,R.W.
Unpublished
JOURNAL 2 (bases 1 to 170682)
AUTHORS Bruno,D., Conn,L., Della Rosa,M., Faulkner,D., Federspiel,N.,
Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D.,
Yu,S. and Davis,R.W.
Direct Submission
TITLE Submitted (14-FEB-2000) DNA Sequencing and Technology Center,
JOURNAL Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT On Apr 10, 2001 this sequence version replaced gi:13562078.
-----Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SBDTC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
-----Project Information
Center project name: 837
Center clone name: Rp11-121M22
-----Summary Statistics
Sequencing Vector: M13mp18; X02513
Chemistry: Dye-primer; 12% of reads
Chemistry: Dye-terminator Big Dye; 86% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165770 bases at least Q40
Consensus quality: 167255 bases at least Q30
Consensus quality: 167918 bases at least Q20
Insert size: 172423; agarose-fp
Quality coverage: 9.0x in Q20 bases; agarose-fp
Quality coverage: 9.1x in Q20 bases; sum-of-contigs.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
```

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1856: contig of 1856 bp in length
 1857 1956: gap of unknown length
 1957 7065: contig of 5109 bp in length
 7066 7165: gap of unknown length
 7166 21089: contig of 13924 bp in length
 21090 21189: gap of unknown length
 21190 34145: contig of 12956 bp in length
 34146 34245: gap of unknown length
 34246 48562: contig of 14317 bp in length
 48563 48662: gap of unknown length
 48663 109179: contig of 60517 bp in length
 109180 109279: gap of unknown length
 109280 170682: contig of 61403 bp in length.

FEATURES

source

1..170682
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP11-121M22"
 /clone_lib="RC1 human BAC library 11"

misc_feature

1..1856
 /note="assembly_name:Contig25"

misc_feature

1957..7065
 /note="assembly_name:Contig26"

misc_feature

7166..21089
 /note="assembly_name:Contig27"

misc_feature

21190..34145
 /note="assembly_name:Contig28"

misc_feature

34246..48562
 /note="assembly_name:Contig29"

misc_feature

48663..109179
 /note="assembly_name:Contig30"

misc_feature

109280..170682
 /note="assembly_name:Contig31"

misc_feature

clone_end:17"
 clone_end:T7"

BASE COUNT 43947 a 39732 c 39514 g 46884 t 605 others

ORIGIN

Query Match

Best Local Similarity 28.1%; Score 803; DB 2; Length 170682;

Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 ACTATTTTTCAGATCAGACATTTTCAGGAGACTTTTACCTACACTGACCGCGGATGC 215

Db 98615 ACTATTTTTCAGATCAGACATTTTCAGGAGACTTTTACCTACACTGACCGCGGATGC 98674

QY 216 TCAGTTCTTGGCTCCGGCTTCTCCACTGAGCATCTGGCGCTCCCGCTCCAGGGGCTCAC 275

Db 98675 TCAGTTCTTGGCTCCGGCTTCTCCACTGAGCATCTGGCGCTCCCGCTCCAGGGGCTCAC 98734

QY 276 CGGGGCTCTTCAGACCTGCGACGCTGCTTCTATCTCGGAGCGTGAACCGCGAGCCGGA 335

Db 98735 CGGGGCTCTTCAGACCTGCGACGCTGCTTCTATCTCGGAGCGTGAACCGCGAGCCGGA 98794

QY 336 CTCGTCGCTGCTGAGCCTGTGGCGGGGGCTCCGCGGAGGCTTGTGCTACCGAGGGCG 395

Db 98795 CTCGTCGCTGCTGAGCCTGTGGCGGGGGCTCCGCGGAGGCTTGTGCTACCGAGGGCG 98854

QY 396 CGAGTATGTCATTAGCCGCTCCCAATGCTAGCGCGCGGGCGGCGCAACAGCA 455

Db 98855 CGAGTATGTCATTAGCCGCTCCCAATGCTAGCGCGCGGGCGGCGCAACAGCA 98914

QY 456 GGGCGACACCTTCTCCAGCGCGGGGTGTTCCGGGGCGGGCTTCCGGAGACCCACCTC 515

Db 98915 GGGCGACACCTTCTCCAGCGCGGGGTGTTCCGGGGCGGGCTTCCGGAGACCCACCTC 98974

QY 516 TCGCTGCGGGTGGCTCGGGCTGGAACCCGCCATCTACGGCCCTTGACCCCTTACAA 575

|||||

Db 98975 TCGCTGCGGGTGGCTCGGGCTGGAACCCGCCATCTTAGGGCCCTGACCCCTTACAA 99034

QY 576 GCCCGGGGGGGGGCTTCGGGGAGAGTCTGTAGCCGGCGAGGCTCTGGGCGCGCAAGCG 635

Db 99035 GCCCGGGGGGGGGCTTCGGGGAGAGTCTGTAGCCGGCGAGGCTCTGGGCGCGCAAGCG 99094

QY 636 TTTCTGTCTATCCCGGGTACGTGGAGAGCGTGGTGTGTCGCGGACGAGTCAATGGTCAA 695

Db 99095 TTTCTGTCTATCCCGGGTACGTGGAGAGCGTGGTGTGTCGCGGACGAGTCAATGGTCAA 99154

QY 696 GTTCCAGCGCGGACCTGGAACATTATCTGTGACGCTCTGSCAACGGCGCGGACT 755

Db 99155 GTTCCAGCGCGGACCTGGAACATTATCTGTGACGCTCTGSCAACGGCGCGGACT 99214

QY 756 CTACCGGCATCCAGCATCTCAACCCCATCAACATCGTTGTGGTCAAGTGTGCTTCT 815

Db 99215 CTACCGGCATCCAGCATCTCAACCCCATCAACATCGTTGTGGTCAAGTGTGCTTCT 99274

QY 816 TAGAGATCGTGACTCCGGGGCCCAAGGTCAAGGCAATCGCGCCCTGACGCTGCGCAACTT 875

Db 99275 TAGAGATCGTGACTCCGGGGCCCAAGGTCAAGGCAATCGCGCCCTGACGCTGCGCAACTT 99334

QY 876 CTGTGCTGCGCAGAGAGCTGAACAAAGTGAACAGAGTGAACAGACCCCGAGTACTGGGACAC 935

Db 99335 CTGTGCTGCGCAGAGAGCTGAACAAAGTGAACAGAGTGAACAGACCCCGAGTACTGGGACAC 99394

QY 936 TGCATCTCTTCCACAGGCGAGG 958

Db 99395 TGCATCTCTTCCACAGGCGAGG 99417

RESULT 7

AP003459 28000 bp DNA linear PRI 08-MAR-2002
 LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:RP11-211H5,
 DEFINITION complete sequence.

ACCESSION AP003459

VERSION AP003459.2 GI:19263045

KEYWORDS HTG.

SOURCE Homo sapiens DNA, clone:RP11-211H5.

ORGANISM Homo sapiens

REFERENCE 1

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens genomic DNA

JOURNAL Published Only in Database (2001)

REFERENCE 2 (bases 1 to 28000)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (29-MAR-2001) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/;

Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT On Mar 7, 2002 this sequence version replaced gi:13488920.

FEATURES Location/Qualifiers

1..28000

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="11q"

/clone="RP11-211H5"

BASE COUNT 6938 a 6083 c 6679 g 8300 t

ORIGIN

Query Match 27.2%; Score 775; DB 9; Length 28000;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2079 GCATGGCTACAATTTCGTGGTGGCCATCCCGCGAGCGGCGCTCAAGCATCGACATCCGCCA 2138

```
|||||
Db 3473 GCATGGCTACAAATTCGTGGTGGCCATCCCGCAGCGCCTCAAGCATCGACATCCGCCA 3532
QY 2139 GCGCGGTACAAAGGCTGATCGGGATGACAACTACCTGCTCTGAAGACACGCCAAGG 2198
Db 3533 GCGCGGTACAAAGGCTGATCGGGATGACAACTACCTGCTCTGAAGACACGCCAAGG 3592
QY 2199 CAAGTACCTGCTCAACGGGCATTTCTGTTGTCGGCGGTGGAGGGGACCTGGTGTGAA 2258
Db 3593 CAAGTACCTGCTCAACGGGCATTTCTGTTGTCGGCGGTGGAGGGGACCTGGTGTGAA 3652
QY 2259 GGGCAGTCTGCTCGGTACAGCGGCACGGGCACAGCGGTGGAGAGCCTGCAAGGCTTCCCG 2318
Db 3653 GGGCAGTCTGCTCGGTACAGCGGCACGGGCACAGCGGTGGAGAGCCTGCAAGGCTTCCCG 3712
QY 2319 GGGCAGTCTGCTCGGTACAGCGGCACGGGCACAGCGGTGGAGAGCCTGCAAGGCTTCCCG 2378
Db 3713 GGGCAGTCTGCTCGGTACAGCGGCACGGGCACAGCGGTGGAGAGCCTGCAAGGCTTCCCG 3772
QY 2379 GGTGCGCTACTCTTCTATCTGCCCAAGAGCGCTCGGAGGACAAAGTCTCTCATCCCAA 2438
Db 3773 GGTGCGCTACTCTTCTATCTGCCCAAGAGCGCTCGGAGGACAAAGTCTCTCATCCCAA 3832
QY 2439 GGACCCCGGGGAGCCCTCTGTCTTGCACACAGCGTCTCTCAGCCCTCTCCAAACAGGTGGA 2498
Db 3833 GGACCCCGGGGAGCCCTCTGTCTTGCACACAGCGTCTCTCAGCCCTCTCCAAACAGGTGGA 3892
QY 2499 GCGCGGGAGCAGACGCCCTCTGACGCTGGTGTGCTGCGAGCTGGGGCGCTGCTCCGC 2558
Db 3893 GCGCGGGAGCAGACGCCCTCTGACGCTGGTGTGCTGCGAGCTGGGGCGCTGCTCCGC 3952
QY 2559 GAGCTGCGGAGTGGCTGAGAGCGGGCGGTGGAGTCCCGGGGCTCCCGCGGCGAGCG 2618
Db 3953 GAGCTGCGGAGTGGCTGAGAGCGGGCGGTGGAGTCCCGGGGCTCCCGCGGCGAGCG 4012
QY 2619 CAGGTCCTCTGCTGTGATGAGCCCTGCGCCCTGTGAGACACAGCCTGCGGGGAGCC 2678
Db 4013 CAGGTCCTCTGCTGTGATGAGCCCTGCGCCCTGTGAGACACAGCCTGCGGGGAGCC 4072
QY 2679 CTGCCCCACCTGGGAGCTCAGCGCTGCTGACCTGCTCCAAAGAGCTGCGGCGGGGATT 2738
Db 4073 CTGCCCCACCTGGGAGCTCAGCGCTGCTGACCTGCTCCAAAGAGCTGCGGCGGGGATT 4132
QY 2739 TCAGAGGCGCTCAGTCAAGTGTGGGACAGCGGCGGCTGCTGCGGCGGACAGTG 2798
Db 4133 TCAGAGGCGCTCAGTCAAGTGTGGGACAGCGGCGGCTGCTGCGGCGGACAGTG 4192
QY 2799 CAAGTGCACCGCAAGCCCGAGAGCTGAGTCTGCTGAGCGCTGCTGAGCGTGTCTGA 2853
Db 4193 CAAGTGCACCGCAAGCCCGAGAGCTGAGTCTGCTGAGCGCTGCTGAGCGTGTCTGA 4247

RESULT 8
AC025130 172905 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 11 clone RP11-211H6 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
AC025130
VERSION AC025130.2 GI:7387384
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172905)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
AUTHORS Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
TITLE Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G., Collins,S.,
JOURNAL Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galan,J., Gargana,S., Ginde,S., Govette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 172905)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
AUTHORS Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
TITLE Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G., Collins,S.,
JOURNAL Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galan,J., Gargana,S., Ginde,S., Govette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2000 this sequence version replaced gi:7158941.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7808
Center clone name: 211_H_6
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16189 bases at least Q40
Consensus quality: 167402 bases at least Q30
Consensus quality: 169801 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 171405; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
```

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 21: contig of 21 bp in length
22 121: gap of 100 bp
122 1343: contig of 1222 bp in length
1344 1443: gap of 100 bp
1444 2636: contig of 1193 bp in length
2637 2736: gap of 100 bp
2737 4052: contig of 1316 bp in length
4053 4152: gap of 100 bp
4153 8044: contig of 3892 bp in length
8045 8144: gap of 100 bp
8145 14949: contig of 6805 bp in length
14950 15049: gap of 100 bp
15050 23535: contig of 8486 bp in length
23536 23635: gap of 100 bp
23636 30261: contig of 6626 bp in length
30262 30361: gap of 100 bp
30362 40941: contig of 10580 bp in length
40942 41041: gap of 100 bp
41042 54893: contig of 13852 bp in length
54894 54993: gap of 100 bp
54994 68888: contig of 13895 bp in length
68889 68988: gap of 100 bp
68989 83702: contig of 14714 bp in length
83703 83802: gap of 100 bp
83803 98393: contig of 14591 bp in length
98394 98493: gap of 100 bp
98494 117356: contig of 18863 bp in length
117357 117456: gap of 100 bp
117457 140589: contig of 23133 bp in length
140590 140690: gap of 100 bp
140690 172905: contig of 32216 bp in length.
  
```

FEATURES

Source

```

1..172905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone_lib="RPC1-11 Human Male BAC"
1..21
/feature="assembly_fragment"
clone_end:77
vector_side:right
122..1343
/feature="assembly_fragment"
1444..2636
/feature="assembly_fragment"
2737..4052
/feature="assembly_fragment"
4153..8044
/feature="assembly_fragment"
8145..14949
/feature="assembly_fragment"
15050..23535
/feature="assembly_fragment"
23636..30261
/feature="assembly_fragment"
clone_end:SP6
vector_side:right
30362..40941
/feature="assembly_fragment"
41042..54893
/feature="assembly_fragment"
54994..68888
/feature="assembly_fragment"
68989..83702
/feature="assembly_fragment"
83803..98393
/feature="assembly_fragment"
  
```

```

misc_feature 98494..117356
/feature="assembly_fragment"
misc_feature 117457..140589
/feature="assembly_fragment"
misc_feature 140590..172905
/feature="assembly_fragment"
BASE COUNT 44360 a 41333 c 39516 g 46194 t 1502 others
ORIGIN
Query Match 12.5%; Score 358; DB 2; Length 172905;
Best Local Similarity 100.0%; Pred. No. 2.1e-183;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2079 GCATGGCTACAAATTCGTGGTGGCCATCCCGCGAGGCGCCTCAAGCATCCGACATCCGCA 2138
Db 54519 GCATGGCTACAAATTCGTGGTGGCCATCCCGCGAGGCGCCTCAAGCATCCGACATCCGCA 54578
QY 2139 GCGCGGTTACAAAGGCTGATCGGGGATGACAACTACTGCTGCTGAAGACACA:CCCAAGG 2198
Db 54579 GCGCGGTTACAAAGGCTGATCGGGGATGACAACTACTGCTGCTGAAGACACA:CCCAAGG 54638
QY 2199 CAAGTACCTGCTCAACGGGCAATTCGTGGTTCGCGGTCGAGCGGACCTGGTGGTGA 2258
Db 54639 CAAGTACCTGCTCAACGGGCAATTCGTGGTTCGCGGTCGAGCGGACCTGGTGGTGA 54698
QY 2259 GGGCAGTCTGCTGGGTCATACGGGACGGGCACAGCGGTGAGAGCCTGAGGCTTCCG 2318
Db 54699 GGGCAGTCTGCTGGGTCATACGGGACGGGCACAGCGGTGAGAGCCTGAGGCTTCCG 54758
QY 2319 GCCATCTCGGAGCGGTGACCGTGGAGGTCTCTCGGTGGGGAAGATGACACGCGCCG 2378
Db 54759 GCCATCTCGGAGCGGTGACCGTGGAGGTCTCTCGGTGGGGAAGATGACACGCGCCG 54818
QY 2379 GGTCCGTACTCTCTCTATCTGCCCCAAGAGCCTCGGAGGACAAAGTCTCTCATCCC 2436
Db 54819 GGTCCGTACTCTCTCTATCTGCCCCAAGAGCCTCGGAGGACAAAGTCTCTCATCCC 54876
  
```

RESULT 9
 AC023429/c

LOCUS
 DEFINITION Homo sapiens chromosome 11 clone RP11-121M22, WORKING DRAFT

ACCESSION
 AC023429

VERSION
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS
 SOURCE Homo sapiens.

ORGANISM
 Homo sapiens

REFERENCE
 1 (bases 1 to 170682)

AUTHORS
 Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D.,

Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,

Mao, J., Komp, C., Kottler, S., Lam, B., Marathe, R., Miranda, M.,

Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D.,

Southwick, A.M., Webb, C., Wilhelmy, J., Yu, S. and Davis, R.W.

Unpublished

JOURNAL
 2 (bases 1 to 170682)

AUTHORS
 Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D., Federspiel, N.,

Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,

Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J.,

Yu, S. and Davis, R.W.

TITLE
 JOURNAL

COMMENT
 On Apr 10, 2001 this sequence version replaced gi:13562078.
 ----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center
 Center code: SDSTDC
 Web site: <http://sequence-www.stanford.edu/group/human/>
 Contact: hum-info@sequence.stanford.edu

```
----- Project Information
Center project name: 837
Center clone name: RP11-121M22
----- Summary Statistics
Sequencing Vector: M13mp18; X02513
Chemistry: Dye-primer; 12% of reads
Chemistry: Dye-terminator Big Dye; 86% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165770 bases at least Q40
Consensus quality: 167255 bases at least Q30
Consensus quality: 167918 bases at least Q20
Insert size: 172423; agarose-fp
Insert size: 170082; sum-of-contigs
Quality coverage: 9.0x in Q20 bases; agarose-fp
Quality coverage: 9.1x in Q20 bases; sum-of-contigs
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1856: contig of 1856 bp in length
* 1857 1956: gap of unknown length
* 1957 7065: contig of 5109 bp in length
* 7066 7165: gap of unknown length
* 7166 21089: contig of 13924 bp in length
* 21090 21189: gap of unknown length
* 21190 34145: contig of 12956 bp in length
* 34146 34246: gap of unknown length
* 34246 48562: contig of 14317 bp in length
* 48563 48663: gap of unknown length
* 48663 109179: contig of 60517 bp in length
* 109180 109279: gap of unknown length
* 109280 170682: contig of 61403 bp in length.
FEATURES
    source
        1..170682
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /clone="RP11-121M22"
            /clone_lib="RPC1 human BAC library 11"
        1..1856
            /note="assembly_name:Contig25"
            1957..7065
            /note="assembly_name:Contig26"
            7166..21089
            /note="assembly_name:Contig27"
            21190..34145
            /note="assembly_name:Contig28"
            34246..48562
            /note="assembly_name:Contig29"
            clone_end:SP6
            48663..109179
            /note="assembly_name:Contig30"
            109280..170682
            /note="assembly_name:Contig31"
            clone_end:77
BASE COUNT 43947 a 39732 c 39514 g 46884 t 605 others
ORIGIN
Query Match 10.1%; Score 288; DB 2; Length 170682;
Best Local Similarity 100.0%; Pred. No. 3.4e-145;
Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1258 GGTGACTGCTCTGTCGACCAACCCAGCCATCTCCCTGCCGAGATCTGCCGGGC 1317
|
|
|
Db 44571 GGTGACTGCTCTGTCGACCAACCCAGCCATCTCCCTGCCGAGATCTGCCGGGC 44512
|
|
|
QY 1318 GCCAGCTACACCTGAGCCAGCAGTCGCGAGCTGGCTTTTGGCTGGGCTCCCAAGCCCTGT 1377
|
|
|
Db 44511 GCCAGCTACACCTGAGCCAGCAGTCGCGAGCTGGCTTTTGGCTGGGCTCCCAAGCCCTGT 44512
```


Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Correll, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Jacobson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Lucier, R., Luna, R., Ma, J.,
 Lozadó, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczkyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 178764)
 Worley, K.C.
 Direct Submission
 Submitted (06-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 178764)
 Worley, K.C.
 Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: G2F0
 Center clone name: CH230-254N12
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.950329
 Consensus quality: 124573 bases at least Q40
 Consensus quality: 131850 bases at least Q30
 Consensus quality: 137700 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 49 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1037: contig of 1037 bp in length
 * 1038 1137: gap of unknown length
 * 1138 2250: contig of 1113 bp in length
 *

2251 gap of unknown length
 2351 contig of 1229 bp in length
 3580 gap of unknown length
 3680 contig of 1195 bp in length
 4875 gap of unknown length
 4975 contig of 1551 bp in length
 6526 gap of unknown length
 6536 contig of 1280 bp in length
 7906 gap of unknown length
 8005 contig of 1719 bp in length
 9074 gap of unknown length
 9824 contig of 1618 bp in length
 9835 contig of 1358 bp in length
 11443 gap of unknown length
 11543 contig of 1150 bp in length
 12901 gap of unknown length
 13001 contig of 1315 bp in length
 14151 gap of unknown length
 14251 contig of 1810 bp in length
 15566 gap of unknown length
 15666 contig of 1115 bp in length
 17476 gap of unknown length
 18691 gap of unknown length
 18791 contig of 1479 bp in length
 20269 gap of unknown length
 20369 contig of 1558 bp in length
 21927 gap of unknown length
 22027 contig of 1971 bp in length
 23998 gap of unknown length
 24098 contig of 1581 bp in length
 25679 gap of unknown length
 25779 contig of 1649 bp in length
 27429 gap of unknown length
 27529 contig of 2149 bp in length
 29678 gap of unknown length
 29778 contig of 1100 bp in length
 30878 gap of unknown length
 30978 contig of 1854 bp in length
 32832 gap of unknown length
 32932 contig of 2488 bp in length
 35420 gap of unknown length
 35519 contig of 1874 bp in length
 37393 gap of unknown length
 37493 contig of 2342 bp in length
 39835 gap of unknown length
 39936 contig of 2854 bp in length
 42890 gap of unknown length
 42899 gap of unknown length
 46647 contig of 3757 bp in length
 46747 gap of unknown length
 49189 contig of 2442 bp in length
 49289 gap of unknown length
 53805 contig of 4517 bp in length
 53905 gap of unknown length
 56202 contig of 2297 bp in length
 56302 gap of unknown length
 58719 contig of 2417 bp in length
 58819 gap of unknown length
 62904 contig of 4085 bp in length
 63004 gap of unknown length
 65932 contig of 2928 bp in length
 66033 gap of unknown length
 70120 contig of 4088 bp in length
 70220 gap of unknown length
 73258 contig of 3038 bp in length
 73358 gap of unknown length
 77466 contig of 4108 bp in length
 77566 gap of unknown length
 80545 contig of 2979 bp in length
 80646 gap of unknown length
 85988 contig of 5342 bp in length
 86087 gap of unknown length
 90908 contig of 4821 bp in length
 91008 gap of unknown length

```

* 91009 97918: contig of 6910 bp in length
* 97919 98018: gap of unknown length
* 98019 103502: contig of 5484 bp in length
* 103503 103602: gap of unknown length
* 103603 112570: contig of 8968 bp in length
* 112571 112670: gap of unknown length
* 112671 117924: contig of 5254 bp in length
* 117925 118024: gap of unknown length
* 118025 125019: contig of 6995 bp in length
* 125020 125119: gap of unknown length
* 125120 134087: contig of 8968 bp in length
* 134088 134187: gap of unknown length
* 134188 143851: contig of 9664 bp in length
* 143852 143951: gap of unknown length
* 143952 151799: contig of 7748 bp in length
* 151799 151799: gap of unknown length
* 151800 158826: contig of 7027 bp in length
* 158827 158926: gap of unknown length
* 158927 168522: contig of 9596 bp in length
* 168523 168623: gap of unknown length
* 168623 178764: contig of 10142 bp in length.
FEATURES
    source
    1. .178764
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-254N12"
BASE COUNT 47487 a 39119 c 38897 g 45423 t 7838 others
ORIGIN
Query Match      2.1%  Score 61;  DB 2;  Length 178764;
Best Local Similarity 100.0%;  Pred. No. 3e-21;  0;  Gaps 0;
Matches 61;  Conservative 0;  Mismatches 0;  Indels 0;

Qy 1942 GTCGTGTCACGAAGTCATCAAGCTGGCTGTGATGGGAACCTGGGCTCCAGAAG 2001
|||||
Db 150003 GTCGTGTCACGAAGTCATCAAGCTGGCTGTGATGGGAACCTGGGCTCCAGAAG 149944

Qy 2002 A 2002
Db 149943 A 149943

RESULT 12
AC126507
LOCUS
DEFINITION Rattus norvegicus clone CH230-254N12, *** SEQUENCING IN PROGRESS.
AC126507
VERSION AC126507.1 GI:21700463
KEYWORDS HTG; HTGS PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 178764)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,

```

```

Homsí, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, F., Johnson, R., Jollivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, C., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morjan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, S.,
Oraguñe, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Peters, L., Pickens, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Neilson, D.,
Weinstock, G. and Gibbs, R.
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZFU
Center clone name: CH230-254N12
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124573 bases at least Q40
Consensus quality: 131850 bases at least Q30
Consensus quality: 137700 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1037: contig of 1037 bp in length
* 1137: gap of unknown length
* 1138 1138: contig of 1113 bp in length
* 2250 2250: gap of unknown length
* 2251 2351: contig of 1229 bp in length
* 3579 3579: gap of unknown length
* 3679 3679: gap of unknown length
* 3680 4874: contig of 1195 bp in length
* 4875 4974: gap of unknown length
* 4975 5525: contig of 1551 bp in length
* 5525 6625: gap of unknown length

```

6626 7905: contig of 1280 bp in length
7906: gap of unknown length
8006 9724: contig of 1719 bp in length
9725 9824: gap of unknown length
9825 11442: contig of 1618 bp in length
11443 11542: gap of unknown length
11543 12900: contig of 1358 bp in length
12901 13000: gap of unknown length
13001 14150: contig of 1150 bp in length
14151 14250: gap of unknown length
14251 15365: contig of 1315 bp in length
15366 15665: gap of unknown length
15666 17475: contig of 1810 bp in length
17476 17575: gap of unknown length
17576 18690: contig of 1115 bp in length
18691 18990: gap of unknown length
18991 20269: contig of 1479 bp in length
20270 20369: gap of unknown length
20370 21927: contig of 1558 bp in length
21928 22027: gap of unknown length
22028 23998: contig of 1971 bp in length
23999 24098: gap of unknown length
24099 25679: contig of 1581 bp in length
25680 25779: gap of unknown length
25780 27428: contig of 1649 bp in length
27429 27528: gap of unknown length
27529 29677: contig of 2149 bp in length
29678 29777: gap of unknown length
29778 30977: contig of 1100 bp in length
30978 30977: gap of unknown length
30979 32831: contig of 1854 bp in length
32832 32931: gap of unknown length
32932 35419: contig of 2488 bp in length
35420 35519: gap of unknown length
35520 37433: contig of 1874 bp in length
37434 37433: gap of unknown length
37435 39835: contig of 2342 bp in length
39836 39935: gap of unknown length
39936 42789: contig of 2854 bp in length
42790 42889: gap of unknown length
42891 46646: contig of 3757 bp in length
46647 46746: gap of unknown length
46747 49188: contig of 2442 bp in length
49189 49288: gap of unknown length
49289 53805: contig of 4517 bp in length
53806 53905: gap of unknown length
53906 56202: contig of 2297 bp in length
56203 56302: gap of unknown length
56303 58719: contig of 2417 bp in length
58720 58819: gap of unknown length
58820 62904: contig of 4085 bp in length
62905 63004: gap of unknown length
63006 65932: contig of 2928 bp in length
65933 66032: gap of unknown length
66033 70130: contig of 4088 bp in length
70131 70220: gap of unknown length
70221 73258: contig of 3038 bp in length
73259 73358: gap of unknown length
73359 77466: contig of 4108 bp in length
77467 77566: gap of unknown length
77567 80545: contig of 2979 bp in length
80546 80645: gap of unknown length
80646 85987: contig of 5342 bp in length
85988 86087: gap of unknown length
86088 90908: contig of 4821 bp in length
90909 91008: gap of unknown length
91010 97918: contig of 6910 bp in length
97919 98018: gap of unknown length
98019 103502: contig of 5484 bp in length
103503 103602: gap of unknown length
103603 112570: contig of 8968 bp in length
112571 112670: gap of unknown length
112671 117924: contig of 5254 bp in length

117925 118024: gap of unknown length
118025 125019: contig of 6995 bp in length
125020 125119: gap of unknown length
125120 134087: contig of 8968 bp in length
134088 134187: gap of unknown length
134188 143851: contig of 9664 bp in length
143852 143951: gap of unknown length
143952 151699: contig of 7748 bp in length
151700 151799: gap of unknown length
151800 158826: contig of 7027 bp in length
158827 158926: gap of unknown length
158927 168522: contig of 9596 bp in length
168523 168622: gap of unknown length
168623 178764: contig of 10142 bp in length.

FEATURES
Location/Qualifiers
1..178764
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-254N12"

BASE COUNT 47487 a 39119 c 38897 g 45423 t 7838 others
ORIGIN

Query Match 1.5% Score 44: DB 2: Length 178764;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 ACCTCTCGTGGGGTGGCTCGGCTGGAACCCCGCATCCT 554
|||||
Db 101958 ACCTCTCGTGGGGTGGCTCGGCTGGAACCCCGCATCCT 102001
|||||

RESULT 13
AC101990/c
LOCUS
DEFINITION Mus musculus clone RP24-371J2, WORKING DRAFT SEQUENCE, 16 unordered
pieces.
AC101990 182656 bp DNA linear HTG 21-AUG-2002
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 182656)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP24-371J2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182656)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,W., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,I., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Melidrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zaloun,J., Zembek,B., Zimmer,A. and Zody,M.

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

REFERENCE
AUTHORS

3 (bases 1 to 182656)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Direct Submission
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17060766.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L17833
 Center clone name: 371-J_2
 ----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 178030 bases at least Q40
 Consensus quality: 179964 bases at least Q30
 Consensus quality: 180647 bases at least Q20
 Insert size: 172000; agarose-ff
 Insert size: 181156; sum-of-contigs
 Quality coverage: 8.1 in Q20 bases; agarose-ff
 Quality coverage: 7.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 361: contig of 361 bp in length
 * 362 461: gap of 100 bp
 * 462 1538: contig of 1077 bp in length
 * 1539 1638: gap of 100 bp
 * 1639 3397: contig of 1759 bp in length
 * 3398 3497: gap of 100 bp
 * 3498 4947: contig of 1450 bp in length
 * 4948 5047: gap of 100 bp
 * 5048 6893: contig of 1846 bp in length
 * 6894 6993: gap of 100 bp
 * 6994 10395: contig of 3402 bp in length
 * 10396 10495: gap of 100 bp
 * 10496 12584: contig of 2089 bp in length
 * 12585 12684: gap of 100 bp
 * 12685 14977: contig of 2293 bp in length
 * 14978 15077: gap of 100 bp
 * 15078 18181: contig of 3104 bp in length
 * 18182 18281: gap of 100 bp
 * 18282 22654: contig of 4373 bp in length

* 22655 22754: gap of 100 bp
 * 22755 31261: contig of 8507 bp in length
 * 31262 31361: gap of 100 bp
 * 31362 43587: contig of 12226 bp in length
 * 43588 43687: gap of 100 bp
 * 43688 57612: contig of 13925 bp in length
 * 57613 57712: gap of 100 bp
 * 57713 77467: contig of 19755 bp in length
 * 77468 77567: gap of 100 bp
 * 77568 108724: contig of 31157 bp in length
 * 108725 108824: gap of 100 bp
 * 108825 182656: contig of 73832 bp in length.

FEATURES

Location/Qualifiers
 1..182656
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP24-371J2"
 /clone_lib="RPC1-24 Male Mouse BAC"
 1..361
 /note="assembly_fragment"

misc_feature

462..1538
 /note="assembly_fragment"

misc_feature

1639..3397
 /note="assembly_fragment"

misc_feature

3498..4947
 /note="assembly_fragment"

misc_feature

5048..6893
 /note="assembly_fragment"

misc_feature

6994..10395
 /note="assembly_fragment"

misc_feature

10496..12584
 /note="assembly_fragment"

misc_feature

12685..14977
 /note="assembly_fragment"

misc_feature

15078..18181
 /note="assembly_fragment"

misc_feature

18282..22654
 /note="assembly_fragment"

misc_feature

22755..31261
 /note="assembly_fragment"

misc_feature

31362..43587
 /note="assembly_fragment"

misc_feature

43688..57612
 /note="assembly_fragment"

misc_feature

57713..77467
 /note="assembly_fragment"

misc_feature

77568..108724
 /note="assembly_fragment"

misc_feature

108825..182656
 /note="assembly_fragment"

BASE COUNT

51468 a 41160 c 39770 g 48758 t 1500 others
 ORIGIN

Query Match 1.3%; Score 36; DB 2; Length 182656;
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 ACAGCCAGGGCGCACACCTTCTCCAGCGCGGGTG 484
 |||||||||||||||||||||||||||||||||||||||

DB 39334 ACAGCCAGGGCGCACACCTTCTCCAGCGCGGGTG 39299

RESULT 14

BC009667

LOCUS

DEFINITION

BC009667

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC009667 3927 bp mRNA linear ROD 07-AUG-2002
 Mus musculus, similar to a disintegrin and metalloproteinase with
 thrombospondin motifs 1 (ADAMTS-1), clone IMAGE:3491991, mRNA,
 partial cds.
 BC009667 GI:16307155
 house mouse.
 Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3927)
Strausberg, R.
Direct Submission
Submitted (29-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRK plate: 8 Row: J Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
Location/Qualifiers
1. 3927
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="CZECH II"
/clone="IMAGE:3491991"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clonal) outgrowth infected with the virus MMV."
/clone_lib="NCI_CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
cl. 1024
/codon_start=2
/product="similar to a disintegrin and metalloproteinase
with thrombospondin motifs 1 (ADAMTS-1)"
/protein_id="AAH09667.1"
/db_xref="GI:16307156"
IKAGDGNLGGKPKFKGCGGDKCKRTGLFTKPMHGYFWFVWALPAGASSIDIR
ORGYKGLIDDDNIALKNSQGYLLGHFVSAVERDLVKGSLVLYSGTCTAVESLQ
ASRPLEPLTEVLSVKGWTPRVYSFYLPKPREDKSFRPKDPRGSLVLRNLSVL
SNQVQPNRRPARKVAGSWGPCSYSGSLQKRAVDCRSPGQGSACVDHRLPLE
KRACEPCTWELGNWSPKSCGGRFKRPLKCVGHGRLLRDQCDLRKRPQELDF
CVLRPC"
BASE COUNT 938 a 920 c 1093 g 976 t
ORIGIN
Query Match 1.2%; Score 35; DB 10; Length 3927;
Best Local Similarity 100.0%; Pred. No. 6.4e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1942 GTCTGTGTCCAGGCAAGTCATCAAGCTGGCTG 1976
|||||
Db 113 GTCTGTGTCCAGGCAAGTCATCAAGCTGGCTG 147
RESULT 15
AC099229
LOCUS
DEFINITION
Rattus norvegicus clone CH230-3019, *** SEQUENCING IN PROGRESS ***,
48 unordered pieces.
AC099229
AC099229.3 GI:21736081
VERSION
KEYWORDS
HTG; HTGS_PHASE1.

SOURCE
ORGANISM

Norway rat.

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 86229)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Avele, M., Banks, T.,
Barbata, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
Buck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chaves, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Devila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,
Honsli, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsone, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE
Direct Submission

Unpublished

2 (bases 1 to 86229)

REFERENCE
AUTHORS

Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 86229)

REFERENCE
AUTHORS

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 12, 2002 this sequence version replaced gi:17973111.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GHDX

Center clone name: CH230-3019

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

*	1	1288:	contig	of 1288	bp in length
*	1289	1388:	gap	of unknown	length
*	1389	2302:	contig	of 1114	bp in length
*	2503	2602:	gap	of unknown	length
*	2603	3645:	contig	of 1043	bp in length
*	3646	3745:	gap	of unknown	length
*	3746	4866:	contig	of 1121	bp in length
*	4867	4966:	gap	of unknown	length
*	4967	6411:	contig	of 1445	bp in length
*	6412	6511:	gap	of unknown	length
*	6512	7670:	contig	of 1159	bp in length
*	7671	7770:	gap	of unknown	length
*	7771	8823:	contig	of 1053	bp in length
*	8824	8923:	gap	of unknown	length
*	8924	10105:	contig	of 1182	bp in length
*	10106	10205:	gap	of unknown	length
*	10206	11283:	contig	of 1078	bp in length
*	11284	11383:	gap	of unknown	length
*	11384	12892:	contig	of 1509	bp in length
*	12893	12992:	gap	of unknown	length
*	12993	15116:	contig	of 2124	bp in length
*	15117	15216:	gap	of unknown	length
*	15217	16395:	contig	of 1179	bp in length
*	16396	16495:	gap	of unknown	length
*	16496	17597:	contig	of 1102	bp in length
*	17598	17697:	gap	of unknown	length
*	17698	18822:	contig	of 1125	bp in length
*	18823	18922:	gap	of unknown	length
*	18923	20128:	contig	of 1206	bp in length
*	20129	20228:	gap	of unknown	length
*	20229	22372:	contig	of 2144	bp in length
*	22373	22472:	gap	of unknown	length
*	22473	23954:	contig	of 1482	bp in length
*	23955	24054:	gap	of unknown	length
*	24055	25745:	contig	of 1691	bp in length
*	25746	25845:	gap	of unknown	length
*	25846	27386:	contig	of 1541	bp in length
*	27387	27486:	gap	of unknown	length
*	27487	28937:	contig	of 1451	bp in length
*	28938	29037:	gap	of unknown	length
*	29038	30061:	contig	of 1024	bp in length
*	30062	30161:	gap	of unknown	length
*	30162	31758:	contig	of 1597	bp in length
*	31759	31858:	gap	of unknown	length
*	31859	33015:	contig	of 1157	bp in length
*	33016	33115:	gap	of unknown	length
*	33116	34536:	contig	of 1421	bp in length
*	34537	34636:	gap	of unknown	length
*	34637	35951:	contig	of 1315	bp in length
*	35952	36051:	gap	of unknown	length
*	36052	37616:	contig	of 1465	bp in length
*	37617	37616:	gap	of unknown	length
*	37617	39220:	contig	of 1604	bp in length
*	39221	39320:	gap	of unknown	length
*	39321	40785:	contig	of 1465	bp in length
*	40786	40885:	gap	of unknown	length
*	40886	42770:	contig	of 1885	bp in length
*	42771	42870:	gap	of unknown	length
*	42871	45007:	contig	of 2137	bp in length

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 22:50:37 ; Search time 606 Seconds
(without alignments)
10602.230 Million cell updates/sec

Title: US-10-009-332-2

Perfect score: 2853

Sequence: 1 atgcttttgctgggcatctc.....gcgtcctgagccgtgctga-2853

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2853	100.0	2853	22	AAH41003 Human metalloprotease
2	2695	94.5	2853	24	AAD35569 Human protease cDN
3	2695	94.5	3446	24	AAD35571 Human protease cDN
4	2089	73.2	2930	24	ABK12894 Human protease PRT
5	2089	73.2	2937	24	AAS7182 Human metalloprotease
6	1035	36.3	1104	24	AAD35570 Human metalloprotease cDN
7	903	31.7	966	24	AAD35568 Human protease cDN
8	481	16.9	1143	21	AAA95831 Human metalloprotease
9	50	1.8	1518	21	AAA95827 Rat metalloprotein

c	10	30	1.1	30	22	AAH41029	Adaptor primer SEQ
c	11	30	1.1	30	22	AAH41030	Adaptor primer SEQ
c	12	30	1.1	41	22	AAH41017	PCR primer specifi
c	13	30	1.1	42	22	AAH41016	PCR primer specifi
c	14	27	0.9	27	22	AAH41018	PCR primer specifi
c	15	27	0.9	27	22	AAH41019	PCR primer specifi
c	16	27	0.9	38	22	AAH41024	PCR primer specifi
c	17	27	0.9	38	22	AAH41033	PCR primer for met
c	18	27	0.9	38	22	AAH41034	PCR primer for met
c	19	25	0.9	610	24	ABQ44966	Oligonucleotide fo
c	20	25	0.9	610	24	ABQ44967	Oligonucleotide fo
c	21	23	0.8	2114	20	AAH19958	Rat Tango-76 encod
c	22	22	0.8	1790	21	AAH48015	Zea mays DNA fragm
c	23	22	0.8	1813	21	AAH48015	Fragment of diamn
c	24	21	0.7	1827	21	AAH48015	Human pancreatic c
c	25	20	0.7	607	23	ABV18267	Human prostate exp
c	26	20	0.7	765	23	ABV18267	Human prostate exp
c	27	20	0.7	1244	21	AAH47449	Arabidopsis thalia
c	28	20	0.7	1247	21	AAH47449	Arabidopsis thalia
c	29	20	0.7	1802	23	ABLO1979	Drosophila melanog
c	30	20	0.7	3802	23	ABLO1979	Drosophila melanog
c	31	20	0.7	5880	22	AAH68687	Pseudomonas putida
c	32	19	0.7	11355	22	AAH68692	Pseudomonas putida
c	33	19	0.7	423	21	AAH95829	Bovine metalloprot
c	34	19	0.7	431	21	AAH95829	Bovine metalloprot
c	35	19	0.7	493	23	AAH86803	Human pancreatic c
c	36	19	0.7	518	23	AAH86803	Human pancreatic c
c	37	19	0.7	1046	22	AAH86803	DNA encoding novel
c	38	19	0.7	1063	24	ABQ25996	CNS disorder-relat
c	39	19	0.7	1063	24	ABQ25996	Human protein HP10
c	40	19	0.7	1063	24	ABQ25996	Oligonucleotide fo
c	41	19	0.7	2089	23	AB199790	Oligonucleotide fo
c	42	19	0.7	2586	23	AB199790	Mouse ischaemic co
c	43	19	0.7	4479	24	ABN60018	Drosophila melanog
c	44	19	0.7	4533	21	AAA53920	Novel human coding
c	45	19	0.7	8139	22	ABA17511	Type III adenylyl
c	46	19	0.7	8139	22	ABA17511	Human nervous syst
c	47	19	0.7	8139	22	ABA17511	Human nervous syst

ALIGNMENTS

RESULT 1
AAH41003
ID AAH41003 standard; cDNA; 2853 BP.
XX
AC AAH41003;
XX
DT 23-AUG-2001 (first entry)
XX
XX Human metalloprotease MDTs6 cDNA.
DE
KW Metalloprotease; human; aggreganase; joint disease; osteoarthritis;
KW osteopathic; antiarthritic; ss.
XX
XX Homo sapiens.
OS
FH Key Location/Qualifiers
CDS 1..2853
FT /*tag- a
FT /product= "Metalloprotease"
FT /note= "Aggreganase activity"
XX
XX WO200134785-A1.
XX
XX PD 17-MAY-2001.
XX
XX PF 10-NOV-2000; 2000WO-JP07917.
XX
XX PR 11-NOV-1999; 99JP-0321740.
XX
XX PR 16-MAY-2000; 2000JP-0144020.
XX
XX PA (YAMA) YAMANOUCHI PHARM CO LTD.
XX (KAZU-) KAZUSA DNA RES INST.

XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
 XX WPI; 2001-343602/36.
 DR P-PSDB; AAG62299.
 XX Metalloprotease with aggrecanase activity for treating joint diseases
 PT especially osteoarthritis
 XX
 XX Example 2; Page 60-61; 85pp; Japanese.
 XX
 CC This invention relates to a metalloprotease with aggrecanase activity.
 CC The invention includes protein and DNA sequences of the metalloprotease,
 CC vectors containing the DNA, host cells transformed by the vectors, and
 CC antibodies directed against the metalloprotease. The antibodies, protein
 CC and DNA sequences can be used in the treatment and prevention of joint
 CC diseases, particularly osteoarthritis. The treatment may result in
 CC osteopathic and antiarthritic activity. The present sequence represents
 CC cDNA encoding the metalloprotease termed MDT56.
 XX
 SQ Sequence 2853 BP; 522 A; 920 C; 903 G; 508 T; 0 other;
 Query Match 100.0%; Score 2853; DB 22; Length 2853;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2853; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCTTTTGTGGGATCCTAACCTGGCTTCGCGGGGGAACCGCTGGAGGCTTTAG 60
 DB 1 ATGCTTTTGTGGGATCCTAACCTGGCTTCGCGGGGGAACCGCTGGAGGCTTTAG 60
 QY 61 CCAGAGCGGAGGTAGTCTCCATCCGACTGGACCCCGGACATTAACGCCGCCGCTAC 120
 DB 61 CCAGAGCGGAGGTAGTCTCCATCCGACTGGACCCCGGACATTAACGCCGCCGCTAC 120
 QY 121 TACTGGCGGGTCCCGAGGACTCCGGGGATCAGGGACTCATTTTTCAGATCAGACATTT 180
 DB 121 TACTGGCGGGTCCCGAGGACTCCGGGGATCAGGGACTCATTTTTCAGATCAGACATTT 180
 QY 181 CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGGCTTCTCC 240
 DB 181 CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGGCTTCTCC 240
 QY 241 ACTGAGATCTGGGCTCCCGTCCAGGGCTCACCGGGGCTTTCAGACCTTGCAGCCG 300
 DB 241 ACTGAGATCTGGGCTCCCGTCCAGGGCTCACCGGGGCTTTCAGACCTTGCAGCCG 300
 QY 301 TGCTTCTATTCTGGGACGTGAACCGCGAGCGGACTCGTTCGCTGCTGAGGCTGTGC 360
 DB 301 TGCTTCTATTCTGGGACGTGAACCGCGAGCGGACTCGTTCGCTGCTGAGGCTGTGC 360
 QY 361 GGGGGTCCCGGAGGCTTTGGCTACCGAGCGCGGAGTATGCTATTAGCCCGCTGCC 420
 DB 361 GGGGGTCCCGGAGGCTTTGGCTACCGAGCGCGGAGTATGCTATTAGCCCGCTGCC 420
 QY 421 AATGCTAGCGCGCGGCGGCGAGCAACAGCCAGCGGCGACACCTTCTCCAGCGCCGG 480
 DB 421 AATGCTAGCGCGCGGCGGCGAGCAACAGCCAGCGGCGACACCTTCTCCAGCGCCGG 480
 QY 481 GGTGTTCCGGCGGGCTTCCGAGAGCCACCTCTCGCTCGGGGTGGGCTCGGGCTGG 540
 DB 481 GGTGTTCCGGCGGGCTTCCGAGAGCCACCTCTCGCTCGGGGTGGGCTCGGGCTGG 540
 QY 541 AACCCCGCATCTACGGGCGCTTGAACCCCTTACAGCCGCGGGGGGGGCTTCGGGGAG 600
 DB 541 AACCCCGCATCTACGGGCGCTTGAACCCCTTACAGCCGCGGGGGGGGCTTCGGGGAG 600
 QY 601 AGTGTAGCGCGCGGCTTGGCGCGGCAAGCGTTTCGTTATCCCGCGGTAGGTG 660
 DB 601 AGTGTAGCGCGCGGCTTGGCGCGGCAAGCGTTTCGTTATCCCGCGGTAGGTG 660
 QY 661 GAGACGCTGGTGGTTCGGGAGGAGTCAATGATCAAGTTCACGGCGCGGACCTTGGACAT 720
 DB 661 GAGACGCTGGTGGTTCGGGAGGAGTCAATGATCAAGTTCACGGCGCGGACCTTGGACAT 720

QY 721 TATCTGTGACGCTGCTGGCAACGCGCGCGGACTCTACCGCCATCCAGCATCCTCAAC 780
 DB 721 TATCTGTGACGCTGCTGGCAACGCGCGCGGACTCTACCGCCATCCAGCATCCTCAAC 780
 QY 781 CCCATCAACATCGTGTGGTCAAGGTGCTTCTTAGAGATCGTGACTCGGGGCCAAG 840
 DB 781 CCCATCAACATCGTGTGGTCAAGGTGCTTCTTAGAGATCGTGACTCGGGGCCAAG 840
 QY 841 GTACCGGCAATGGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAGAACTGAAC 900
 DB 841 GTACCGGCAATGGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAGAACTGAAC 900
 QY 901 AAAGTGTGACGAGCAACCCGAGTACTGGGACACTGCGCATCTCTTACAGGAGAGAC 960
 DB 901 AAAGTGTGACGAGCAACCCGAGTACTGGGACACTGCGCATCTCTTACAGGAGAGAC 960
 QY 961 CTGTGTGAGGCGGACCACTGTGACACCTGGCATGGTGATGGGTACATCTGTGAC 1020
 DB 961 CTGTGTGAGGCGGACCACTGTGACACCTGGCATGGGTGATGGGTACATCTGTGAC 1020
 QY 1021 CCCAAGAGAACTGCTCTGTCTATTGAGGACGATGGGCTTCCATCAGCCTTCCACTGCC 1080
 DB 1021 CCCAAGAGAACTGCTCTGTCTATTGAGGACGATGGGCTTCCATCAGCCTTCCACTGCC 1080
 QY 1081 CACGAGCTGGGCGACGTGTTCAACATGCCCATGACAAATGTGAAAGTCTGTGAGAGGTG 1140
 DB 1081 CACGAGCTGGGCGACGTGTTCAACATGCCCATGACAAATGTGAAAGTCTGTGAGAGGTG 1140
 QY 1141 TTTTGGAGAGCTCCGAGCAACCATGATGTCGCCGACCTTCATCCAGATCGACCGTGCC 1200
 DB 1141 TTTTGGAGAGCTCCGAGCAACCATGATGTCGCCGACCTTCATCCAGATCGACCGTGCC 1200
 QY 1201 AACCCCTGCTCAGCCTGCGAGTGCCTCATCATCAGCCTTCTTGACAGCGGACCGT 1260
 DB 1201 AACCCCTGCTCAGCCTGCGAGTGCCTCATCATCAGCCTTCTTGACAGCGGACCGT 1260
 QY 1261 GACTGCTCTGGGCAACCCAGCAAGCCCATCTCCCTGCGCGAGGATCTGCCGGGCGCC 1320
 DB 1261 GACTGCTCTGGGCAACCCAGCAAGCCCATCTCCCTGCGCGAGGATCTGCCGGGCGCC 1320
 QY 1321 AGCTACACCTCAGCAGCAGTGCAGTGGCTTTTGGCGTGGGCTCCAAAGCTCTGTCT 1380
 DB 1321 AGCTACACCTCAGCAGCAGTGCAGTGGCTTTTGGCGTGGGCTCCAAAGCTCTGTCT 1380
 QY 1381 TACATGCACTATGACCAAGCTGTGTGACCGGGAAGGCAAGGACACATGTTGTGC 1440
 DB 1381 TACATGCACTATGACCAAGCTGTGTGACCGGGAAGGCAAGGACACATGTTGTGC 1440
 QY 1441 CAGACCGGCACTTCCCTGGGCGGATGGCAGCTGTGGGAGGCAAGCTCTGCCCTC 1500
 DB 1441 CAGACCGGCACTTCCCTGGGCGGATGGCAGCTGTGGGAGGCAAGCTCTGCCCTC 1500
 QY 1501 AAAGGGGCTGCGTGGAGAGACACAACTTCAACAGCAGAGGTGGATGTTCTTGGGCC 1560
 DB 1501 AAAGGGGCTGCGTGGAGAGACACAACTTCAACAGCAGAGGTGGATGTTCTTGGGCC 1560
 QY 1561 AAATGGGATCCCTATGGCCCTGCTGCGCACATGTGTGGGGGCTGCACCTCGCCAGG 1620
 DB 1561 AAATGGGATCCCTATGGCCCTGCTGCGCACATGTGTGGGGGCTGCAGTGGCCAGG 1620
 QY 1621 AGCGAGTGCACCAACCCACCCCTGCAACGGGGCAAGTACTGCGAGGAGTGAAGGTG 1680
 DB 1621 AGCGAGTGCACCAACCCACCCCTGCAACGGGGCAAGTACTGCGAGGAGTGAAGGTG 1680
 QY 1681 AAATACCGATCTGCAACCTGGAGCCCTGCCCCAGCTCAGCCTCGGGAAGAGCTTCCGG 1740
 DB 1681 AAATACCGATCTGCAACCTGGAGCCCTGCCCCAGCTCAGCCTCGGGAAGAGCTTCCGG 1740
 QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACACCAAGGCTGCGGAGGCTCTCTCGCC 1800
 DB 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACACCAAGGCTGCGGAGGCTCTCTCGCC 1800

1801 GTGGCATGGGTGCCAAGTACTCCGGGTGTCTCCCGGGACAAGTGAAGCTCATCTGC 1860
Db GTGGCATGGGTGCCAAGTACTCCGGGTGTCTCCCGGGACAAGTGAAGCTCATCTGC 1860
QY 1861 CGAGCAATGGCACTGGCTACTTCTATGTGCTGGGACACCAAGGTGGTGGACGGCAGGTG 1920
Db 1861 CGAGCAATGGCACTGGCTACTTCTATGTGCTGGGACACCAAGGTGGTGGACGGCAGGTG 1920
QY 1921 TGCTCTCTGACTTCCACCTCCGCTCTGTGTCACCAAGCAAGTGCATCAAGCTGGCTGTAT 1980
Db 1921 TGCTCTCTGACTTCCACCTCCGCTCTGTGTCACCAAGCAAGTGCATCAAGCTGGCTGTAT 1980
QY 1981 GGAACCTGGGTCCCAAGCAAGATTCGACAAAGTGTGGGTGTGGGGGACACAATAAG 2040
Db 1981 GGAACCTGGGTCCCAAGCAAGATTCGACAAAGTGTGGGTGTGGGGGACACAATAAG 2040
QY 2041 AGCTGCAAGAGGTGACTGAGCTCTTCCACCAAGCCATGCTGCTACAAATTCGTGTG 2100
Db 2041 AGCTGCAAGAGGTGACTGAGCTCTTCCACCAAGCCATGCTGCTACAAATTCGTGTG 2100
QY 2101 GCATCCCCCGAGGGCCCTCAAGCATCGACATCCGCCAGCGGTTACAAAGGGGTGATC 2160
Db 2101 GCATCCCCCGAGGGCCCTCAAGCATCGACATCCGCCAGCGGTTACAAAGGGGTGATC 2160
QY 2161 GGGATGACAACTACTTGGCTCTGAAGACACAGCAAGCAAGTACTGCTCAAGGGCAT 2220
Db 2161 GGGATGACAACTACTTGGCTCTGAAGACACAGCAAGCAAGTACTGCTCAAGGGCAT 2220
QY 2221 TTCTGGTGTCTCGCGGTGGAGCGGACCTGTGGTGAAGGGCAGTCTCGCGGTACAGC 2280
Db 2221 TTCTGGTGTCTCGCGGTGGAGCGGACCTGTGGTGAAGGGCAGTCTCGCGGTACAGC 2280
QY 2281 GGCAGGGCACAGCGGTGGAGAGCTGCAAGCTTCCCGGCCCATCTCGAGCCGCTGACC 2340
Db 2281 GGCAGGGCACAGCGGTGGAGAGCTGCAAGCTTCCCGGCCCATCTCGAGCCGCTGACC 2340
QY 2341 GTGGAGTCTCTCTCGTGGGGAAGATGACACGCGCCCGGTGCTGCTCTCTTCTATCTG 2400
Db 2341 GTGGAGTCTCTCTCGTGGGGAAGATGACACGCGCCCGGTGCTGCTCTCTTCTATCTG 2400
QY 2401 CCCAAGAGCTCGGGAGGACAACTCTCTATCCCAAGGACCCCGGGACCTCTCTGTC 2460
Db 2401 CCCAAGAGCTCGGGAGGACAACTCTCTATCCCAAGGACCCCGGGACCTCTCTGTC 2460
QY 2461 TTGCAACACAGCTCTCAGCTCTCCACCAAGTGGAGAGCGCGGAGCAGAGGCCCT 2520
Db 2461 TTGCAACACAGCTCTCAGCTCTCCACCAAGTGGAGAGCGCGGAGCAGAGGCCCT 2520
QY 2521 GCACGCTGGGTGGCTGGCAGCTGGGGGCGCTGCTCCGCGAGCTGCGGAGTGGCTGAG 2580
Db 2521 GCACGCTGGGTGGCTGGCAGCTGGGGGCGCTGCTCCGCGAGCTGCGGAGTGGCTGAG 2580
QY 2581 AAGCGGGGCTGGACTCGCGGGGTTCGCGGGGAGCGAGCGGCTGCTGCTGATGCA 2640
Db 2581 AAGCGGGGCTGGACTCGCGGGGTTCGCGGGGAGCGAGCGGCTGCTGCTGATGCA 2640
QY 2641 GCCATCGCGCTGGAGACACAGCTTGGCGGGAGCGCTCCCGGAGCTGGGAGCTGAGC 2700
Db 2641 GCCATCGCGCTGGAGACACAGCTTGGCGGGAGCGCTCCCGGAGCTGGGAGCTGAGC 2700
QY 2701 GCCTGGTCAACCTGCTCCCAAGAGCTGCGCGCGGGGATTTCAGAGGCGCTCACTCAAGTGT 2760
Db 2701 GCCTGGTCAACCTGCTCCCAAGAGCTGCGCGCGGGGATTTCAGAGGCGCTCACTCAAGTGT 2760
QY 2761 GTGGGCCACGAGGCGGCTGCTGGCGGGGAGCAGTGCACCTTGCACCGCAAGCCCGAG 2820
Db 2761 GTGGGCCACGAGGCGGCTGCTGGCGGGGAGCAGTGCACCTTGCACCGCAAGCCCGAG 2820
QY 2821 GAGCTGGACTTCTGGTCTCTGAGCGCTGCTGA 2853
Db 2821 GAGCTGGACTTCTGGTCTCTGAGCGCTGCTGA 2853

RESULT 2
AAD35569
ID AAD35569 standard; cDNA; 2853 BP.
XX
AC AAD35569;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease cDNA #2.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2853
FT FT /*tag= a
FT FT /product= "Human protease #2"
XX
PN W0200226949-A2.
XX
PD 04-APR-2002.
XX
XX 27-SEP-2001; 2001WO-US30350.
PF
XX 29-SEP-2000; 2000US-236689P.
PR
XX (LEXI-) LEXICON GENETICS INC.
PA
XX Friddle CJ, Hilbun E;
PI
XX WPI; 2002-372123/40.
DR
XX P-PSDB; AAD22541.
XX
XX Novel nucleic acid encoding a human protease, useful as a hybridization
XX probe for screening libraries and assessing gene expression patterns -
XX
XX Claim 1; Page 35-36; 41pp; English.
XX
XX The present sequence is a cDNA encoding novel human protein (NHP),
XX human protease. NHPs share structural similarity with animal proteases
XX particularly zinc metalloproteases. Sequences of the invention are
XX useful in therapeutic, diagnostic and pharmacogenomic applications.
XX NHP polynucleotides are used as hybridization probes for screening
XX libraries and assessing gene expression patterns. They can also be
XX used for treating related biological disorders such as obesity, high
XX blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy.
XX
XX Sequence 2853 BP; 522 A; 920 C; 903 G; 508 T; 0 other;
XX
Query Match 94.5%; Score 2695; DB 24; Length 2853;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2795; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 57 TGAGCCAGAGCGGGAGGTAGTCTCCATCCGACTGGAGCCGACATTAAACGGCGCG 116
Db 57 TGAGCCAGAGCGGGAGGTAGTCTCCATCCGACTGGAGCCGACATTAAACGGCGCG 116
QY 117 CTACTACTGGCGGGTCCGAGGACTCCGGGGATCAGGACTCATTTTTCNGATACAGC 176
Db 117 CTACTACTGGCGGGTCCGAGGACTCCGGGGATCAGGACTCATTTTTCAGATACAGC 176
QY 177 ATTTCAGGAGGACTTTTACTACACCTGAGCGCGGATGCTCAGTCTTGGCTT 236
Db 177 ATTTCAGGAGGACTTTTACTACACCTGAGCGCGGATGCTCAGTCTTGGCTT 236
QY 237 CTCCTGAGCATCTGGGGTCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACTGG 296
Db 237 CTCCTGAGCATCTGGGGTCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACTGG 296

Qy	297	ACGCTGCTTCTATTCTTGGGACGTTGAAGCCGACCGGACCTCGTTTCGCTGCTGTGAGCCT	356
Db	297	ACGCTGCTTCTATTCTTGGGAGCGTGAACCGCGAGCCGGAATCGTTGCGTCTGTGAGCCT	356
Qy	357	GTCCGGGGGCTCCCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCTATTAGCCCGCT	416
Db	357	GTCCGGGGGCTCCCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCTATTAGCCCGCT	416
Qy	417	GCCCAATGCTAGCGCGCGCGCGCGACGCGAACAGCGCAGGGCGACACCTTCTCCAGCG	476
Db	417	GCCCAATGCTAGCGCGCGCGCGCGACGCGAACAGCGCAGGGCGACACCTTCTCCAGCG	476
Qy	477	CCGGGTGTTCCGGGGGCGCTTCCGGAGAGCCACCTCTCTGCTGGGGGTGCGCTCGG	536
Db	477	CCGGGTGTTCCGGGGGCGCTTCCGGAGAGCCACCTCTCTGCTGGGGGTGCGCTCGG	536
Qy	537	CTGGAACCCCGCATCTACGGGGCCTTGACCTCTTACAAGCCGGCGCGGGCTTCGG	596
Db	537	CTGGAACCCCGCATCTACGGGGCCTTGACCTCTTACAAGCCGGCGCGGGCTTCGG	596
Qy	597	GGAGAGTGTAGCCGGCGAGCTTGGCGGGCGCCAAAGGTTTCGTGTCTATCCCGGGTA	656
Db	597	GGAGAGTGTAGCCGGCGAGCTTGGCGGGCGCCAAAGGTTTCGTGTCTATCCCGGGTA	656
Qy	657	CGTGAGAGCCTGGTGGTCCGGAGAGTCAATGGTCAAGTTCCACGGCGCGGACCTGGA	716
Db	657	CGTGAGAGCCTGGTGGTCCGGAGAGTCAATGGTCAAGTTCCACGGCGCGGACCTGGA	716
Qy	717	ACATTATCTGCTAGCGCTGCTGGCAACGGCGGGGAGCTCTACCGCCATCCGAGCATCCT	776
Db	717	ACATTATCTGCTAGCGCTGCTGGCAACGGCGGGGAGCTCTACCGCCATCCGAGCATCCT	776
Qy	777	CAACCCCATCAACATCGTTGGTCAAGGTGCTGCTCTTAGAGATCGTGACTCCCGGCC	836
Db	777	CAACCCCATCAACATCGTTGGTCAAGGTGCTGCTCTTAGAGATCGTGACTCCCGGCC	836
Qy	837	CAAGGTACCGGCAATGCGGCCCTGACGTGCGCAACTTCTGTGCTTGGCAGAGAAGCT	896
Db	837	CAAGGTACCGGCAATGCGGCCCTGACGTGCGCAACTTCTGTGCTTGGCAGAGAAGCT	896
Qy	897	GAACAAGTGAGTGACAACACCCGAGTACTTGGGACACTGCCATCTCTTCCACGAGCA	956
Db	897	GAACAAGTGAGTGACAACACCCGAGTACTTGGGACACTGCCATCTCTTCCACGAGCA	956
Qy	957	GGACCTGTGTGGACCAACCACTGTGACACCTTGGGCATGGCTGATGTGGGTACCATGTG	1016
Db	957	GGACCTGTGTGGACCAACCACTGTGACACCTTGGGCATGGCTGATGTGGGTACCATGTG	1016
Qy	1017	TGACCCCAAGAGAAGCTGCTGTCAATTGAGAGCATGGCTTCCATCAGCCTTCCACC	1076
Db	1017	TGACCCCAAGAGAAGCTGCTGTCAATTGAGAGCATGGCTTCCATCAGCCTTCCACC	1076
Qy	1077	TGCCACGAGCTGGCGACGTGTTCAACATGCCCATGACAATGTGAAGTCTGTGAGGA	1136
Db	1077	TGCCACGAGCTGGCGACGTGTTCAACATGCCCATGACAATGTGAAGTCTGTGAGGA	1136
Qy	1137	GGTGTGGAAGCTCGAGCCACACATGATGTCCCGGACCTCATCCAGATCGACCG	1196
Db	1137	GGTGTGGAAGCTCGAGCCACACATGATGTCCCGGACCTCATCCAGATCGACCG	1196
Qy	1197	TGCCAACCCCTGTGAGCTGAGTGCTGCCATCATCAGCACTTCTTGGACACGGGCA	1256
Db	1197	TGCCAACCCCTGTGAGCTGAGTGCTGCCATCATCAGCACTTCTTGGACACGGGCA	1256
Qy	1257	CGGTGACTTGCTTGGACCAACCCAGCAAGCCCATCTCCCTGCCCCGAGGATCTGCCGG	1316
Db	1257	CGGTGACTTGCTTGGACCAACCCAGCAAGCCCATCTCCCTGCCCCGAGGATCTGCCGG	1316
Qy	1317	CGCCAGCTACACCTGAGCCAGAGTGCGAGCTGGCTTTGGCGTGGGCTCCAGAGCCCTG	1376
Db	1317	CGCCAGCTACACCTGAGCCAGAGTGCGAGCTGGCTTTGGCGTGGGCTCCAGAGCCCTG	1376
Qy	1377	TCCTTACATCAGTACTGCAACCAAGCTGTGGTGCACCCGGGAAGGCCAAGGACAGATGGT	1436

Db	1377	TCCTTTACATGCAGTACTGCACCAAGCTGTGTGTCACCGGGAAGGCCAAGGCACAGATGGT	1433
Qy	1437	GTGCCAGACCCGCCACTTCCCTGGGCCGATGGCACCAAGCTGTGTGGCGAGGGCAAGCTCTGT	1496
Db	1437	GTGCCAGACCCGCCACTTCCCTGGGCCGATGGCACCAAGCTGTGTGGCGAGGGCAAGCTCTGT	1496
Qy	1497	CCTCAAGAGGGCCCTGCGTGGAGAGACACAACCTCAACAGCACAGGTGTGATGCTTCCCTG	1556
Db	1497	CCTCAAGAGGGCCCTGCGTGGAGAGACACAACCTCAACAGCACAGGTGTGATGCTTCCCTG	1556
Qy	1557	GGCCAAATGGGATCCCTATATGGCCCTGTCTCGCGCACATGTGTGTGGGGCGTGCAGCTGGC	1616
Db	1557	GGCCAAATGGGATCCCTATATGGCCCTGTCTCGCGCACATGTGTGTGGGGCGTGCAGCTGGC	1616
Qy	1617	CAGGAGCAGTGCACCAACCCACCCCTGCCAACCGGGGGCAAGTACTCCGAGGAGAGTGAG	1676
Db	1617	CAGGAGCAGTGCACCAACCCACCCCTGCCAACCGGGGGCAAGTACTCCGAGGAGAGTGAG	1676
Qy	1677	GGTGAATACCGATCCTGTCAACCTGAGAGCCCTGCCAACCGGGGGCAAGTACTCCGAGGAGAGCTT	1736
Db	1677	GGTGAATACCGATCCTGTCAACCTGAGAGCCCTGCCAACCGGGGGCAAGTACTCCGAGGAGAGCTT	1736
Qy	1737	CCGGGAGGACGATGTGAGGCTTTCAACCGCTTACAACACACAGCACACCAGCCGCTCACTCT	1796
Db	1737	CCGGGAGGACGATGTGAGGCTTTCAACCGCTTACAACACACAGCACACCAGCCGCTCACTCT	1796
Qy	1797	CGCCGTGGCATGGGTGCCCAAGTACTCCCGCGTGTCTCCCGGGACAAAGTGCAGAGCTCAT	1856
Db	1797	CGCCGTGGCATGGGTGCCCAAGTACTCCCGCGTGTCTCCCGGGACAAAGTGCAGAGCTCAT	1856
Qy	1857	CTGCCAGCCAAATGGCACTGGCTTCTATGTGTGGCACCCAAAGTGTGGACGGGCAC	1916
Db	1857	CTGCCAGCCAAATGGCACTGGCTTCTATGTGTGGCACCCAAAGTGTGGACGGGCAC	1916
Qy	1917	GCTGTGCTCTCTGACTCCACCTCCCTCTGTGTCAAGCAAGTGCATCAAGGCTGGCGTG	1976
Db	1917	GCTGTGCTCTCTGACTCCACCTCCCTCTGTGTCAAGCAAGTGCATCAAGGCTGGCGTG	1976
Qy	1977	TGATGGGAACCTGGGTCTCCAAGAAGAGATTGCACAAGTGTGGGTGTGTGGGGGAGACAA	2036
Db	1977	TGATGGGAACCTGGGTCTCCAAGAAGAGATTGCACAAGTGTGGGTGTGTGGGGGAGACAA	2036
Qy	2037	TAAGAGCTGAAGAAGTGACTGAGCTCTTCAACAGCCCATGCATGCTACAACTTCCTGT	2096
Db	2037	TAAGAGCTGAAGAAGTGACTGAGCTCTTCAACAGCCCATGCATGCTACAACTTCCTGT	2096
Qy	2097	GGTGGCCATCCCCGACAGGCGCCTCAAGCATCGACATCGACATCGCCACAGCGCGTTACAAAGGCT	2156
Db	2097	GGTGGCCATCCCCGACAGGCGCCTCAAGCATCGACATCGACATCGCCACAGCGCGTTACAAAGGCT	2156
Qy	2157	GATCGGGATGACAACCTACCTGGCTCTGAAGACACAGCAAGCAAGTACTGCTCAACGG	2216
Db	2157	GATCGGGATGACAACCTACCTGGCTCTGAAGACACAGCAAGCAAGTACTGCTCAACGG	2216
Qy	2217	GCATTTCTGTGTCTCGCGTGGAGGGGACCTGTGTGTGAAGGGCAGTCTGCTCGCGTA	2276
Db	2217	GCATTTCTGTGTCTCGCGTGGAGCGGACCTGTGTGTGAAGGGCAGTCTGCTCGCGTA	2276
Qy	2277	CAGCGGCACGGGCACACAGGCTGGAGAGCCTGTCCGGCCCATCTCTGGAGCCGCT	2336
Db	2277	CAGCGGCACGGGCACACAGGCTGGAGAGCCTGTCCGGCCCATCTCTGGAGCCGCT	2336
Qy	2337	GACCTGTGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGCTCCGCTACTCTCTCTA	2396
Db	2337	GACCTGTGAGTCTCTCTCCGTGGGGAAGATGACACCGCCCGGCTCCGCTACTCTCTCTA	2396
Qy	2397	TCTGCCCCAAAGAGCCTCGGAGGACCAAGTCTCTCATCCAAAGACCCCGGGGACCCCTC	2456
Db	2397	TCTGCCCCAAAGAGCCTCGGAGGACCAAGTCTCTCATCCAAAGACCCCGGGGACCCCTC	2456
Qy	2457	TGCTTTGCACAACAGCGTCTCAGCCCTCTCCACCAAGTGTGAGCAGCCGCGACAGAGCC	2516
Db	2457	TGCTTTGCACAACAGCGTCTCAGCCCTCTCCACCAAGTGTGAGCAGCCGCGACAGAGCC	2516

Db 2457 TGTCTGCACACAGCGTCTCTAGCCCTCTCCAAACCAGGTGGAGCAGCGGACGAGGCC 2516
 QY 2517 CCCTGCACAGCTGGGTGCTGGCAGCTGGGGCGGTGCTCGCGAGAGTGGGAGTGGCCT 2576
 Db 2517 CCCTGCACAGCTGGGTGCTGGCAGCTGGGGCGGTGCTCGCGAGAGTGGGAGTGGCCT 2576
 QY 2577 GCAGAACGGCGGTGACTGCGGGGCTCCGCCGGGACGCGACCGTCCCTGCTGTGA 2636
 Db 2577 GCAGAACGGCGGTGACTGCTGGGGCTCCGCCGGGACGCGACCGTCCCTGCTGTGA 2636
 QY 2637 TGACGCCCATGCGCCGCTGAGACACAAAGCTCGGGGGAGCCCTGCCACCTGGGAGCT 2696
 Db 2637 TGACGCCCATGCGCCGCTGAGACACAAAGCTCGGGGGAGCCCTGCCACCTGGGAGCT 2696
 QY 2697 CAGCGCTGCTACCCCTGCTCCAAAGAGCTCGGGCGGGGATTTTCAGAGCGCTCACTGAA 2756
 Db 2697 CAGCGCTGCTACCCCTGCTCCAAAGAGCTCGGGCGGGGATTTTCAGAGCGCTCACTGAA 2756
 QY 2757 GTGTGTGGCCACGGAGCGCGGTGCTGGCCGGGACCAAGTGCACCGCAAGCC 2816
 Db 2757 GTGTGTGGCCACGGAGCGCGGTGCTGGCCGGGACCAAGTGCACCGCAAGCC 2816
 QY 2817 CCAGGAGCTGACTTCTGCGTCTCTGAGGCGGTGCTGA 2853
 Db 2817 CCAGGAGCTGACTTCTGCGTCTCTGAGGCGGTGCTGA 2853

RESULT 3
 AAD35571
 ID AAD35571 standard; cDNA; 3446 BP.
 AC AAD35571;
 XX
 DT 26-JUL-2002 (first entry)
 DE Human protease cDNA #4.
 XX
 KW Human; novel human protein; NHP; protease; biological disorder; obesity;
 KW high blood pressure; arthritis; connective tissue disorder; infertility;
 KW gene therapy; enzyme; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200226949-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 27-SEP-2001; 2001WO-US30350.
 XX
 PR 29-SEP-2000; 2000US-236689P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Friddle CJ, Hilbun B;
 XX
 DR WPI; 2002-372123/40.
 XX
 PT Novel nucleic acid encoding a human protease, useful as a hybridization
 PT probe for screening libraries and assessing gene expression patterns -
 XX
 PS Disclosure; Page 40-41; 41pp; English.
 XX
 CC The present sequence is a cDNA encoding novel human protein (NHP),
 CC human protease. NHPs share structural similarity with animal proteases
 CC particularly zinc metalloproteases. Sequences of the invention are
 CC useful in therapeutic, diagnostic and pharmacogenomic applications.
 CC NHP polynucleotides are used as hybridization probes for screening
 CC libraries and assessing gene expression patterns. They can also be
 CC used for treating related biological disorders such as obesity, high
 CC blood pressure, arthritis, connective tissue disorders and infertility.
 CC They are also used in gene therapy.
 XX
 SQ Sequence 3446 BP; 612 A; 1114 C; 1101 G; 619 T; 0 other;

Query Match 94.5%; Score 2695; DB 24; Length 3446;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2795; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 57 TGAGCCACAGCGGGAGGTAGTCTCCCATCGAGCTGACCGGACATTAACGGCCGCCG 116
 Db 453 TGAGCCACAGCGGGAGGTAGTCTCCCATCGAGCTGACCGGACATTAACGGCCGCCG 512
 QY 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCAATTTTTCAGATCAGC 176
 Db 513 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCAATTTTTCAGATCAGC 572
 QY 177 ATTTTCAGGAGGACTTTTACCTACACCTGACCGCGGATCTCAGTCTCTGGCTCCGCCCT 236
 Db 573 ATTTTCAGGAGGACTTTTACCTACACCTGACCGCGGATCTCAGTCTCTGGCTCCGCCCT 632
 QY 237 CTCCACTGAGCATCTGGGCGTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTCGC 296
 Db 633 CTCCACTGAGCATCTGGGCGTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTCGC 692
 QY 297 AGCTGCTTCTATTCTGGGAGCTGAACCGGAGCTCGTTTCGCTGCTGTGAGCCT 356
 Db 693 AGCTGCTTCTATTCTGGGAGCTGAACCGGAGCTCGTTTCGCTGCTGTGAGCCT 752
 QY 357 GTGCGGGGGCTTCGCGGAGGCTTTTGGCTACCGGGGCTCACCGGGGCTCTTCAGACCTCGC 416
 Db 753 GTGCGGGGGCTTCGCGGAGGCTTTTGGCTACCGGGGCTCACCGGGGCTCTTCAGACCTCGC 812
 QY 417 GCCCAATGCTAGCGCGCGGGGCGGACGCAACAGCGGCGGCGCACACCTTCTTCAGCG 476
 Db 813 GCCCAATGCTAGCGCGCGGGGCGGACGCAACAGCGGCGGCGCACACCTTCTTCAGCG 872
 QY 477 CCGGGGTTTCCGGGCGGGCTTCGCGGAGACCCACCTCTCGCTGCGGGTGGGCTCGGG 932
 Db 873 CCGGGGTTTCCGGGCGGGCTTCGCGGAGACCCACCTCTCGCTGCGGGTGGGCTCGGG 992
 QY 537 CTGGAACCCCGCATCTTACCGGCGGCTTACAGCGCGGCGGCGGCGGCTTCGG 596
 Db 933 CTGGAACCCCGCATCTTACCGGCGGCTTACAGCGCGGCGGCGGCGGCTTCGG 992
 QY 597 GGAGAGTCTAGCGCGGCGGCTCTGGGCGGCGGCAAGCGTTTCGCTCTATCCCGCGGTA 656
 Db 993 GGAGAGTCTAGCGCGGCGGCTCTGGGCGGCGGCAAGCGTTTCGCTCTATCCCGCGGTA 1052
 QY 657 CGTGAGAGCTGTGTGCTGGGAGGAGTCAATGGTCAAGTTCACGCGGCGGCGGAGCTGGA 716
 Db 1053 CGTGAGAGCTGTGTGCTGGGAGGAGTCAATGGTCAAGTTCACGCGGCGGCGGAGCTGGA 1112
 QY 717 ACATTATCTGTGAGCTGTGCGCAACGCGGCGGCGGCTCTACCGGCTCCAGCATCCT 776
 Db 1113 ACATTATCTGTGAGCTGTGCGCAACGCGGCGGCGGCTCTACCGGCTCCAGCATCCT 1172
 QY 777 CAACCCCATCAACATCGTTGGTCAAGGTGCTCTTTAGAGATCGTGACTCCGGGCC 836
 Db 1173 CAACCCCATCAACATCGTTGGTCAAGGTGCTCTTTAGAGATCGTGACTCCGGGCC 1232
 QY 837 CAAGGTACCGGCAATCGGCGCTGACGCTGCGCAACTCTGTGCTGCGGAGGAGAGCT 896
 Db 1233 CAAGGTACCGGCAATCGGCGCTGACGCTGCGCAACTCTGTGCTGCGGAGGAGAGCT 1292
 QY 897 GAACAAAGTGTAGTGACAAAGCACCCGAGTACTGGGACACTGCCATCTCTTTCACGAGCA 956
 Db 1293 GAACAAAGTGTAGTGACAAAGCACCCGAGTACTGGGACACTGCCATCTCTTTCACGAGCA 1352
 QY 957 GGAGCTGTGGAGGCCACCACTGTGACACCCCTGGGATGGCTGATGGTGGTACATGTG 1016
 Db 1353 GGAGCTGTGGAGGCCACCACTGTGACACCCCTGGGATGGCTGATGGTGGTACATGTG 1412
 QY 1017 TGACCCCAAGAGAGCTGCTCTGCTCATTTAGGAGGAGTGGCTTCCATCAGCTTACCAC 1076
 Db 1413 TGACCCCAAGAGAGCTGCTCTGCTCATTTAGGAGGAGTGGCTTCCATCAGCTTACCAC 1472

```
QY 1077 TGCCACGAGCTGGGCCAGCTGTTCACATGCCCATGACAAATGTGAAGTCTGTGAGGA 1136
DB 1473 TGCCACGAGCTGGGCCAGCTGTTCACATGCCCATGACAAATGTGAAGTCTGTGAGGA 1532
QY 1137 GGTGTTGGGAAGCTCCGAGCAACACATGATGTCCCGAGCCCTCATCCAGATCGACCG 1196
DB 1533 GGTGTTGGGAAGCTCCGAGCAACACATGATGTCCCGAGCCCTCATCCAGATCGACCG 1592
QY 1197 TGCCAAACCCCTGGTCAAGCTGCAGTGTGCCATCATCCAGTACTTCCCTGGACAGCGGCA 1256
DB 1593 TGCCAAACCCCTGGTCAAGCTGCAGTGTGCCATCATCCAGTACTTCCCTGGACAGCGGCA 1652
QY 1257 CGGTGACTGCTCTCTGGACCAACCCAGCAAGCCATCTCCCTGCCCGAGGATCTGCCGGG 1316
DB 1653 CGGTGACTGCTCTCTGGACCAACCCAGCAAGCCATCTCCCTGCCCGAGGATCTGCCGGG 1712
QY 1317 CGCCAGCTACACCTGAGCCAGAGTGCAGCTGGCTTTTGGCTGGGCTCCAAGCCCTG 1376
DB 1713 CGCCAGCTACACCTGAGCCAGAGTGCAGCTGGCTTTTGGCTGGGCTCCAAGCCCTG 1772
QY 1377 TCCTTTACATGCACTACTGCACCAAGCTGTGTGCACCGGAAGGCCAAGGACAGATGGT 1436
DB 1773 TCCTTTACATGCACTACTGCACCAAGCTGTGTGCACCGGAAGGCCAAGGACAGATGGT 1832
QY 1437 GTCCAGACCCGCCACTTCCCTTGGGCCGATGGCACAGCTGTGGCGAGGGCAAGCTCG 1496
DB 1833 GTCCAGACCCGCCACTTCCCTTGGGCCGATGGCACAGCTGTGGCGAGGGCAAGCTCG 1892
QY 1497 CCTCAAGGGGCTGGTGGAGAGACACACCTCAACAGCACAGGTGGATGGTTCCTG 1556
DB 1893 CCTCAAGGGGCTGGTGGAGAGACACACCTCAACAGCACAGGTGGATGGTTCCTG 1952
QY 1557 GGCCAAATGGATCCCTATATGGCCCTGCTCGGCACATGTGTGGGGGCGTGACAGTGGC 1616
DB 1953 GGCCAAATGGATCCCTATATGGCCCTGCTCGGCACATGTGTGGGGGCGTGACAGTGGC 2012
QY 1617 CAGAGGCACTGACCAACCCACCCCTGCCAALGGGGCAAGTACTGCGAGGGAGTGAG 1676
DB 2013 CAGAGGCACTGACCAACCCACCCCTGCCAALGGGGCAAGTACTGCGAGGGAGTGAG 2072
QY 1677 GGTGAATACCGATCTGCAACCTGGAGCCCTGCCCGAGCTCAGCCCTCCGGAAAGAGCTT 1736
DB 2073 GGTGAATACCGATCTGCAACCTGGAGCCCTGCCCGAGCTCAGCCCTCCGGAAAGAGCTT 2132
QY 1737 CCGGGAGGAGCTGTGAGGCTTTTCAAGCGCTTACAACGAGTACAACAGCACCAACCGGCTCACTCT 1796
DB 2133 CCGGGAGGAGCTGTGAGGCTTTTCAAGCGCTTACAACGAGTACAACAGCACCAACCGGCTCACTCT 2192
QY 1797 CGCGTGGCATGGTGGCCCAAGTACTCCGCGGTGTCTCCCGGCAAGTGAAGCTCAT 1856
DB 2193 CGCGTGGCATGGTGGCCCAAGTACTCCGCGGTGTCTCCCGGCAAGTGAAGCTCAT 2252
QY 1857 CTGCGAGCAATGGCACTGGCTACTTCTATGTGCTGGCACCAAGGTGTGGACGGCAC 1916
DB 2253 CTGCGAGCAATGGCACTGGCTACTTCTATGTGCTGGCACCAAGGTGTGGACGGCAC 2312
QY 1917 GCTGTGCTCTCTGACTCCACCTCCGCTGTGTCCAAAGGCAAGTGCATCAAGGCTGGCTG 1976
DB 2313 GCTGTGCTCTCTGACTCCACCTCCGCTGTGTCCAAAGGCAAGTGCATCAAGGCTGGCTG 2372
QY 1977 TGATGGAACTGGGCTCCAGAGAGATTCGACAGTGTGGGCTGTGGGGGACACAA 2036
DB 2373 TGATGGAACTGGGCTCCAGAGAGATTCGACAGTGTGGGCTGTGGGGGACACAA 2432
QY 2037 TAAGAGCTCAAGAGTGTGACTGGCTTCTCAACAGCCCATCATGGCTTACAAATTCGT 2096
DB 2433 TAAGAGCTCAAGAGTGTGACTGGCTTCTCAACAGCCCATCATGGCTTACAAATTCGT 2492
QY 2097 GGTGGCCATCCCGCAGGCGCCTCAAGCATTCGACATCCCGCAGCGGCTTACAAAGGCT 2156
DB 2493 GGTGGCCATCCCGCAGGCGCCTCAAGCATTCGACATCCCGCAGCGGCTTACAAAGGCT 2552
QY 2157 GATCGGGATGACAACTACCTGGCTCTGAGAGACAGCCCAAGTACCTGCTCAACGG 2216
```

```
DB 2553 GATCGGGATGACAACTACCTGGCTCTGAAGAACAGCAAGCAAGTACCTGCTCAACGG 2612
QY 2217 GCATTTGCTGTGTCGCGGTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGCGGTA 2276
DB 2613 GCATTTGCTGTGTCGCGGTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGCGGTA 2672
QY 2277 CAGCGGACCGGGACAGCGGTGGAGAGCTTCGAGGCTTCCCGGCCCATCTCGAGCGCT 2336
DB 2673 CAGCGGACCGGGACAGCGGTGGAGAGCTTCGAGGCTTCCCGGCCCATCTCGAGCGCT 2732
QY 2337 GACCGTGGAGTCTCTCCGTGGGGAAGATGACACCCGCCCGGGTCCGCTACTCTTCTA 2396
DB 2733 GACCGTGGAGTCTCTCCGTGGGGAAGATGACACCCGCCCGGGTCCGCTACTCTTCTA 2792
QY 2397 TCTGCCAAAGAGCTCGGGAGGAAAGTCTCTCATCCAAAGACCCCGGGGACCCCTC 2456
DB 2793 TCTGCCAAAGAGCTCGGGAGGAAAGTCTCTCATCCAAAGACCCCGGGGACCCCTC 2852
QY 2457 TGTCTTGACAAACAGCGTCTCAGCCCTCTCAACAGAGTGGAGCGCGGACGACAGGCC 2516
DB 2853 TGTCTTGACAAACAGCGTCTCAGCCCTCTCAACAGAGTGGAGCGCGGACGACAGGCC 2912
QY 2517 CCCTGACAGCTGGTGGCTGGCAGCTGGGGCCGCTGCTCCGAGCTGCGCAGTGGCCCT 2576
DB 2913 CCCTGACAGCTGGTGGCTGGCAGCTGGGGCCGCTGCTCCGAGCTGCGCAGTGGCCCT 2972
QY 2577 GCAGAAAGCGGCGGTGGACTCGCGGGGCTCCCGCGGCGACGACGCTCCCTGCTGTGA 2636
DB 2973 GCAGAAAGCGGCGGTGGACTCGCGGGGCTCCCGCGGCGACGACGCTCCCTGCTGTGA 3032
QY 2637 TGCAGCCATCGGCCGTGGAGACACAGCTGCGGGGAGCCCTGCCACCTGGGAGCT 2696
DB 3033 TGCAGCCATCGGCCGTGGAGACACAGCTGCGGGGAGCCCTGCCACCTGGGAGCT 3092
QY 2697 CAGGCGCTGGTCAACCTGCTCAAGAGCTGCGCGCGGGGATTCAGAGGCGCTCACTCAA 2756
DB 3093 CAGGCGCTGGTCAACCTGCTCAAGAGCTGCGCGCGGGGATTCAGAGGCGCTCACTCAA 3152
QY 2757 GTGTGTGGCCACGAGGCGCGCTGCTGGCCCGGGACAGTGCACCTTGACCCCAAGCC 2816
DB 3153 GTGTGTGGCCACGAGGCGCGCTGCTGGCCCGGGACAGTGCACCTTGACCCCAAGCC 3212
QY 2817 CCAGAGCTGGACTTCTCGCTCCTGAGGCGCTGTGA 2853
DB 3213 CCAGAGCTGGACTTCTCGCTCCTGAGGCGCTGTGA 3249

RESULT 4
ABK12894
ID ABK12894 standard; cdna; 2930 BP.
XX
AC ABK12894;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRS-11 cDNA sequence.
XX
KW Human; protease; PRS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 75..2930
FT /*tag= a
FT /partial
FT /product= "Human protease PRS-11"
FT /note= "This sequence lacks a stop codon"
```

XX PN WO200198468-A2.
XX PD 27-DEC-2001.
XX PF 13-JUN-2001; 2001WO-US19178.
XX PR 16-JUN-2000; 2000US-212336P.
XX PR 22-JUN-2000; 2000US-213955P.
XX PR 29-JUN-2000; 2000US-215396P.
XX PR 07-JUL-2000; 2000US-216821P.
XX PR 14-JUL-2000; 2000US-218946P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Yue H. Elliott VS. Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Deleagane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Wallia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX DR WPI: 2002-090437/12.
XX DR P-PSDB: AAU74751.
XX PT Twenty one human proteases (referred to as PRS-1 to PRS-21), useful
XX PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
XX PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
XX PT proliferative (e.g. cancer) disorders -
XX PS Claim 5; Page 168; 177pp; English.
XX CC The present invention relates to twenty one new human proteases,
XX CC referred to as PRS-1 to PRS-21. The PRS polynucleotides and
XX CC polypeptides of the invention are useful in the diagnosis, treatment and
XX CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
XX CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
XX CC myocardial infarction, autoimmune/inflammatory e.g. acquired
XX CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
XX CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
XX CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
XX CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
XX CC endometriosis disorders. Numerous other examples of each disorder are
XX CC given in the specification. The present nucleic acid sequence encodes
XX CC the human protease PRS-11 protein of the invention.
XX SQ Sequence 2930 BP; 529 A; 958 C; 931 G; 512 T; 0 other;

Query Match 73.2%; Score 2089; DB 24; Length 2930;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2379; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 57 TGAGCCAGAGCGGAGGTAGTCTCCATCCGACTGGACCCGGACATTAAAGCGCGCG 116
DB 131 TGAGCCAGAGCGGAGGTAGTCTCCATCCGACTGGACCCGGACATTAAAGCGCGCG 190
QY 117 CTACTACTGCGGGGTCGCGAGACTCCGGGATCAGGACATCTTTTCAGATCAGAC 176
DB 191 CTACTACTGCGGGGTCGCGAGACTCCGGGATCAGGACATCTTTTCAGATCAGAC 250
QY 177 ATTTCAGGAGGACTTTTACCTACACTGACCGCGGATGCTCAGTTCTGGCTCCGCGCTT 236
DB 251 ATTTCAGGAGGACTTTTACCTACACTGACCGCGGATGCTCAGTTCTGGCTCCGCGCTT 310
QY 237 CTCCTACTGAGCATCTGGGCGTCCCTCCAGGGGTCACCGGGGCTCTTCAGACCTGCG 296
DB 311 CTCCTACTGAGCATCTGGGCGTCCCTCCAGGGGTCACCGGGGCTCTTCAGACCTGCG 370
QY 297 AGCTGCTTCTATCTGGGAGCTGACCCGCGGACTCGTTCGCTGCTGAGGCT 356
DB 371 AGCTGCTTCTATCTGGGAGCTGACCCGCGGACTCGTTCGCTGCTGAGGCT 430
QY 357 GTGCGGGGGCTCCGCGGAGCTTTGGCTTACCGAGGCGCGAGTATGTCATTAGCCGCT 416
|||||

DB 431 GTGCGGGGGCTCCGCGGAGCCTTTGGCTACCCAGGCGCGAGTATGTCAATTAGCCCGCT 490
QY 417 GCCCAATGCTAGCGCGCGCGCGCGAGCCCAACAGCCAGGCGGCACACTTCTTCCAGGG 476
DB 491 GCCCAATGCTAGCGCGCGCGCGCGAGCCCAACAGCCAGGCGGCACACTTCTTCCAGGG 550
QY 477 CCGGGGTGTTCCGGGGCGGCGCTTCCGGAGAGACCCACCTCTCGCTCGGGGGTGGCTCGGG 536
DB 551 CCGGGGTGTTCCGGGGCGGCGCTTCCGGAGAGACCCACCTCTCGCTCGGGGGTGGCTCGGG 610
QY 537 CTGGAACCCCGCATCTCTACGCGGCTTACAGCCGCGCGGCGCGGGCTTGGG 596
DB 611 CTGGAACCCCGCATCTCTACGCGGCTTACAGCCGCGCGGCGGGCTTGGG 670
QY 597 GGAGAGTCTAGCTAGCGCGCGAGGTCTGGGCGCGCCAAAGGCTTTCTGTGTATATCCCGCGTA 656
DB 671 GGAGAGTCTAGCTAGCGCGCGAGGTCTGGGCGCGCCAAAGGCTTTCTGTGTATATCCCGCGTA 730
QY 657 CCGTGGAGAGCTGTTGGTTCGCGGAGAGTCAATGTTCAAGTTCACGCGGCGGAGCTGGA 716
DB 731 CCGTGGAGAGCTGTTGGTTCGCGGAGAGTCAATGTTCAAGTTCACGCGGCGGAGCTGGA 790
QY 717 ACATTATCTGCTAGCGCTGCTGCGCAACGCGCGGCGGACTCTACCGCCATCCCGAGCATCT 776
DB 791 ACATTATCTGCTAGCGCTGCTGCGCAACGCGCGGCGGACTCTACCGCCATCCCGAGCATCT 850
QY 777 CAACCCCATCAACATCGTTGTTGTTCAAGTTCGTTCTTAGAGATCGTGACTCCGGGCC 836
DB 851 CAACCCCATCAACATCGTTGTTGTTCAAGTTCGTTCTTAGAGATCGTGACTCCGGGCC 910
QY 837 CAAGTTCACCGCAATCGGCGCTGCGCAACGCGCGGCGGACTCTACCGCCATCCCGAGCATCT 896
DB 911 CAAGTTCACCGCAATCGGCGCTGCGCAACGCGCGGCGGACTCTACCGCCATCCCGAGCATCT 970
QY 897 GAACAAAGTGTAGTGAACAGCACCCGAGTACTGGGACACTGCCATCTCTTCAACAGCA 956
DB 971 GAACAAAGTGTAGTGAACAGCACCCGAGTACTGGGACACTGCCATCTCTTCAACAGCA 1030
QY 957 GGACCTGTGTGGAGCCACCACTGTGACACCTGGGCGATGCTGATGTGGGTACCATGTG 1016
DB 1031 GGACCTGTGTGGAGCCACCACTGTGACACCTGGGCGATGCTGATGTGGGTACCATGTG 1090
QY 1017 TGACCCCAAGAGAGCTCTCTGTCTATTGAGGAGCATGGGCTTCATCAGCTTTCACAC 1076
DB 1091 TGACCCCAAGAGAGCTCTCTGTCTATTGAGGAGCATGGGCTTCATCAGCTTTCACAC 1150
QY 1077 TGCCCAACAGCTGGCGCGAGCTGTTCAACATGCCCATGCAATGTGAAGTCTCTGGAGGA 1136
DB 1151 TGCCCAACAGCTGGCGCGAGCTGTTCAACATGCCCATGCAATGTGAAGTCTCTGGAGGA 1210
QY 1137 GGTGTTTGGGAGGCTCCCGAGCAACACATGATGTCCCGGACCTCATCCAGATCGAGCG 1196
DB 1211 GGTGTTTGGGAGGCTCCCGAGCAACACATGATGTCCCGGACCTCATCCAGATCGAGCG 1270
QY 1197 TGCCCAACCTTGGTTCAGCTTCAGTGTGCGCATCATCAGGACTTCTTGGAGTGGGCA 1256
DB 1271 TGCCCAACCTTGGTTCAGCTTCAGTGTGCGCATCATCAGGACTTCTTGGAGTGGGCA 1330
QY 1257 CGGTGACTGCTCTTGGAGCAACCCAGGAGCCCATCTCCCTGCGCGAGATCTGCGCGG 1316
DB 1331 CGGTGACTGCTCTTGGAGCAACCCAGGAGCCCATCTCCCTGCGCGAGATCTGCGCGG 1390
QY 1317 CGCCAGTGTACACCTTGAGCCAGAGTGGCTGGCTTTTGGGCTGGGCTCCAGAGCCCTG 1376
DB 1391 CGCCAGTGTACACCTTGAGCCAGAGTGGGCTGGCTTTTGGGCTGGGCTCCAGAGCCCTG 1450
QY 1377 TCTTTACATGAGTACTGCAACAGCTGTGGTGCACCGGAGGCCAAGGAGGAGATGGT 1436
DB 1451 TCTTTACATGAGTACTGCAACAGCTGTGGTGCACCGGAGGCCAAGGAGGAGATGGT 1510
QY 1437 GTGCCAGAGCCGCGACTTCCCTCGGCGGATGGCAGCAGCTGTGGGAGGCGGAGCTCTG 1496
DB 1511 GTGCCAGAGCCGCGACTTCCCTCGGCGGATGGCAGCAGCTGTGGGAGGCGGAGCTCTG 1570

QY 237 CTCACCTGAGCATCTGGGGTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGG 296
Db 312 CTCACCTGAGCATCTGGGGTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGG 371
QY 297 ACCTGCTTCTATTCTGGGAGAGTGAACCCCGAGCCGGACTCTGCTGCTGTGTGAGCT 356
Db 372 ACCTGCTTCTATTCTGGGAGAGTGAACCCCGAGCCGGACTCTGCTGCTGTGTGAGCT 431
QY 357 GTCCGGGGGCTCCCGGGAGCTTTGGCTACGAGGCGCCGAGTAGTAGTATGATAGCCCGT 416
Db 432 GTCCGGGGGCTCCCGGGAGCTTTGGCTACGAGGCGCCGAGTAGTAGTATGATAGCCCGT 491
QY 417 GCGCAATGTAGCGCGCGCGCGCGAGCGCAACAGCCAGGCGCACACCTTCTCCAGCG 476
Db 492 GCGCAATGTAGCGCGCGCGCGCGAGCGCAACAGCCAGGCGCACACCTTCTCCAGCG 551
QY 477 CCGGGGTGTTCCGGGGGCGCTTCCGGAGACCCCACTCTCGCTCGGGGTGGCTCGG 536
Db 552 CCGGGGTGTTCCGGGGGCGCTTCCGGAGACCCCACTCTCGCTCGGGGTGGCTCGG 611
QY 537 CTGGAACCCCGCATCTACGGGCGCTTGACCCCTTACAAGCGCGCGCGGGGCTTCGG 596
Db 612 CTGGAACCCCGCATCTACGGGCGCTTGACCCCTTACAAGCGCGCGGGGCTTCGG 671
QY 597 GGAGAGTGTAGCGCGCGAGCTGTGGCGGCGCAAGCTTTCTGTCTATCCCGCGGTA 656
Db 672 GGAGAGTGTAGCGCGCGAGCTGTGGCGGCGCAAGCTTTCTGTCTATCCCGCGGTA 731
QY 657 CTTGGAGAGCTGTGTGGCGGAGCTCAATGCTCAAGTTCACGCGCGCGGAGCTTGG 716
Db 732 CTTGGAGAGCTGTGTGGCGGAGCTCAATGCTCAAGTTCACGCGCGGAGCTTGG 791
QY 717 ACATTATCTGTGACGCTGTGGCAACCGCGCGGAGCTTACCGCCATCCACGATCCT 776
Db 792 ACATTATCTGTGACGCTGTGGCAACCGCGCGGAGCTTACCGCCATCCACGATCCT 851
QY 777 CAACCCCATCAACATCTGTGTCAAGTGTCAAGTTCACGCGCGGAGCTTACCGG 836
Db 852 CAACCCCATCAACATCTGTGTCAAGTGTCAAGTTCACGCGCGGAGCTTACCGG 911
QY 837 CAAGTGTACCGGCAATCTGTGACACCTTGGGATGCTGCTTCTAGAGATCTGAGT 896
Db 912 CAAGTGTACCGGCAATCTGTGACACCTTGGGATGCTGCTTCTAGAGATCTGAGT 971
QY 897 GAACAACTGAGTGAACAGCACCGGAGTACTGGGACACTGCCATCTCTTCAACAGGCA 956
Db 972 GAACAACTGAGTGAACAGCACCGGAGTACTGGGACACTGCCATCTCTTCAACAGGCA 1031
QY 957 GGACCTGTGTGGAGCACCACTGTGACACCTTGGGATGCTGCTTCTAGAGATCTGAGT 1016
Db 1032 GGACCTGTGTGGAGCACCACTGTGACACCTTGGGATGCTGCTTCTAGAGATCTGAGT 1091
QY 1017 TGACCCCAAGAGAGCTGCTGTCTGATGAGGAGTGGGCTTCCATCAGCCTTCAACAC 1076
Db 1092 TGACCCCAAGAGAGCTGCTGTCTGATGAGGAGTGGGCTTCCATCAGCCTTCAACAC 1151
QY 1077 TGCCCAAGAGAGCTGCTGTCTGATGAGGAGTGGGCTTCCATCAGCCTTCAACAC 1136
Db 1152 TGCCCAAGAGAGCTGCTGTCTGATGAGGAGTGGGCTTCCATCAGCCTTCAACAC 1211
QY 1137 GGTGTTTGGGAGTTCGAGCAACCACTGATGTCCCGGAGCTTCCATCAGCCTTCAACAC 1196
Db 1212 GGTGTTTGGGAGTTCGAGCAACCACTGATGTCCCGGAGCTTCCATCAGCCTTCAACAC 1271
QY 1197 TGCCCAAGAGAGCTGCTGTCTGATGAGGAGTGGGCTTCCATCAGCCTTCAACAC 1256
Db 1272 TGCCCAAGAGAGCTGCTGTCTGATGAGGAGTGGGCTTCCATCAGCCTTCAACAC 1331
QY 1257 CGGTGACTGCTCTCTGGACCAACCACTGATGTCCCGGAGCTTCCATCAGCCTTCAACAC 1316
Db 1332 CGGTGACTGCTCTCTGGACCAACCACTGATGTCCCGGAGCTTCCATCAGCCTTCAACAC 1391
QY 1317 GCGCAGCTACACCTGTAGCCAGCAGTGCAGCTGGCTTTTGGGCTTCCAGCCCTG 1376

Db 1392 CGCCAGCTTACACCTTGAGCCAGCAGTGCAGCTGGCTTTTGGCTGAGCTTCCAGCCCTG 1451
QY 1377 TCCCTTACATGAGTACTGCAACCAAGCTGTGTGACCGGGAAGGCCAAGGACAGATGGT 1436
Db 1452 TCCCTTACATGAGTACTGCAACCAAGCTGTGTGACCGGGAAGGCCAAGGACAGATGGT 1511
QY 1437 GTGCCAGACCGCCACTTCCCTTGGCCGATGTCACCAAGCTGTGTGAGGAGGCAAGCTCTG 1496
Db 1512 GTGCCAGACCGCCACTTCCCTTGGCCGATGTCACCAAGCTGTGTGAGGAGGCAAGCTCTG 1571
QY 1497 CCTCAAGGGGCGCTGCGTGGAGACACACCTTCAACAGCAGGAGGCTGTGATGGTTCCTG 1556
Db 1572 CCTCAAGGGGCGCTGCGTGGAGACACACCTTCAACAGCAGGAGGCTGTGATGGTTCCTG 1631
QY 1557 GGCCAATGGGATCCCTATGCGCCCTGCTGCGCAGCTGTGTGGGGCTGCGAGCTGGC 1616
Db 1632 GGCCAATGGGATCCCTATGCGCCCTGCTGCGCAGCTGTGTGGGGCTGCGAGCTGGC 1691
QY 1617 CAGGAGCAGTGCACCAACCCCTGCAAGCGGGGCAAGTACTTGCAGGAGTGTGAG 1676
Db 1692 CAGGAGCAGTGCACCAACCCCTGCAAGCGGGGCAAGTACTTGCAGGAGTGTGAG 1751
QY 1677 GGTGAATACCGATCTGCAACCTTGGAGCCCTGCCCCAGCTCAGCCTCCCGAAGAGCTT 1736
Db 1752 GGTGAATACCGATCTGCAACCTTGGAGCCCTGCCCCAGCTCAGCCTCCCGAAGAGCTT 1811
QY 1737 CCGGAGGAGCAGTGTGAGGCTTTCAACGGCTTCAACAGCAGCAGCAACCGCTTCACTCT 1796
Db 1812 CCGGAGGAGCAGTGTGAGGCTTTCAACGGCTTCAACAGCAGCAGCAACCGCTTCACTCT 1871
QY 1797 CCGGATGGGATGGTGGCCCAAGTACTCCGGGCTGTCTCCCGGAGCAAGTGCAGCTCAT 1856
Db 1872 CCGGATGGGATGGTGGCCCAAGTACTCCGGGCTGTCTCCCGGAGCAAGTGCAGCTCAT 1931
QY 1857 CTCCGAGGCAATGGCACTGGCTTCTATGTGCTGGCACCCAA ---GGTGGAGGACGG 1913
Db 1932 CTCCGAGGCAATGGCACTGGCTTCTATGTGCTGGCACCCAAAGTGGT---GGAGCG 1991
QY 1914 CAGCTGTGCTCTCTGACTCCAGCTCCGCTGTGTGCTTCCAGGCAAGTGCATCAGGCTGG 1973
Db 1992 CAGCTGTGCTCTCTGACTCCAGCTCCGCTGTGTGCTTCCAGGCAAGTGCATCAGGCTGG 2051
QY 1974 CTGTATGGGAACCTGGGCTTCCAGAGAGAGATTCGACAGTGTGGGCTGTGTGGGGGAGA 2033
Db 2052 CTGTATGGGAACCTGGGCTTCCAGAGAGATTCGACAGTGTGGGCTGTGTGGGGGAGA 2111
QY 2034 CAATAAGAGTGAAGAAGTGTGACTTCTTCAACAGCCCATCATGCTGGCTTCAATTT 2093
Db 2112 CAATAAGAGTGAAGAAGTGTGACTTCTTCAACAGCCCATCATGCTGGCTTCAATTT 2171
QY 2094 CGTGTGGCCATCCCGCAGGCGCTCAAGCATCGACATCCGCCAGCGGGTTTCAAGAG 2153
Db 2172 CGTGTGGCCATCCCGCAGGCGCTCAAGCATCGACATCCGCCAGCGGGTTTCAAGAG 2231
QY 2154 GCTGATCGGGATGACAACCTTACCTGCTCTGAAGAACAGCCAGCAAGTACTGCTCAA 2213
Db 2232 GCTGATCGGGATGACAACCTTACCTGCTCTGAAGAACAGCCAGCAAGTACTGCTCAA 2291
QY 2214 CGGGCATTTTCTGGTGTCTCGGCGGTGAGGCGGACCTGTGTGAGGCGAGTCTGCTGG 2273
Db 2292 CGGGCATTTTCTGGTGTCTCGGCGGTGAGGCGGACCTGTGTGAGGCGAGTCTGCTGG 2351
QY 2274 GTACAGCGCACGGGACAGCGGTGGAGAGCTTCCCGGCCCATCTCTGGAGCC 2333
Db 2352 GTACAGCGCACGGGACAGCGGTGGAGAGCTTCCCGGCCCATCTCTGGAGCC 2411
QY 2334 GCTGACCGTGGAGTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCCTACTCTCT 2393
Db 2412 GCTGACCGTGGAGTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCCTACTCTCT 2471
QY 2394 CTATCTGCCCAAGAGCTTCCGGAGGAGCAAGTCTCTCTATCCC 2436

```
Db 2472 CTATCTGCCAAAGAGCCTCGGAGGACAAAGTCTCTCATCCC 2514
RESULT 6
AAD35570
ID AAD35570 standard; cDNA; 1104 BP.
XX
AC AAD35570;
XX
XX 26-JUL-2002 (first entry)
XX
DE Human protease cDNA #3.
XX
XX Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX 1..1104
XX /tag= a
XX /product= "Human protease #3"
XX
XX WO200226949-A2.
XX
XX 04-APR-2002..
XX
XX 27-SEP-2001; 2001WO-US30350.
XX
XX 29-SEP-2000; 2000US-236689P.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Friddle CJ, Hilbun E;
XX
XX WPI; 2002-372123/40.
XX
XX P-PSDB; AAE22542.
XX
XX Novel nucleic acid encoding a human protease, useful as a hybridization
XX probe for screening libraries and assessing gene expression patterns -
XX
XX Disclosure; Page 38-39; 41pp; English.
XX
XX The present sequence is a cDNA encoding novel human protein (NHP),
XX human protease. NHPs share structural similarity with animal proteases
XX particularly zinc metalloproteases. Sequences of the invention are
XX useful in therapeutic, diagnostic and pharmacogenomic applications.
XX NHP polynucleotides are used as hybridisation probes for screening
XX libraries and assessing gene expression patterns. They can also be
XX used for treating related biological disorders such as obesity, high
XX blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy..
XX
XX Sequence 1104 BP; 189 A; 357 C; 345 G; 213 T; 0 other;
XX
XX Query Match 36.3%; Score 1035; DB 24; Length 1104;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1035; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Oy 57 TGAGCCAGAGCGGAGGTAGTCTCCATCCGACTGGACCCGACATTAAACGGCCGCG 116
Db 57 TGAGCCAGAGCGGAGGTAGTCTCCATCCGACTGGACCCGACATTAAACGGCCGCG 116
Oy 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGGATCAGGACTCATTTTTCAGATCACAGC 176
Db 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGGATCAGGACTCATTTTTCAGATCACAGC 176
Oy 177 ATTTTCAGAGGACTTTTACCTACACCTGACCCGGATCTCAGTCTTGGCTCCCGCCTT 236
Db 177 ATTTTCAGAGGACTTTTACCTACACCTGACCCGGATCTCAGTCTTGGCTCCCGCCTT 236
Oy 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTACCCGGGGCTCTTCAGACCTGCG 296
```

XX OS Homo sapiens.
 XX PH Key Location/Qualifiers
 XX FT CDS 1..966
 XX FT /*tag= a
 XX FT /product= "Human protease #1"
 XX PN WO200226949-A2.
 XX PN 04-APR-2002.
 XX PF 27-SEP-2001; 2001WO-US30350.
 XX PR 29-SEP-2000; 2000US-236689P.
 XX PA (LEXI-) LEXICON GENETICS INC.
 XX PI Friddle CJ, Hilbun E;
 XX DR WPI; 2002-372123/40.
 XX DR P-PSDB; AAE22540.
 XX PT Nucleic acid encoding a human protease, useful as a hybridization
 XX PT probe for screening libraries and assessing gene expression patterns -
 XX PS Disclosure; Page 34; 41pp; English.
 XX CC The present sequence is a cDNA encoding novel human protein (NHP),
 CC human protease. NHPs share structural similarity with animal proteases
 CC particularly zinc metalloproteases. Sequences of the invention are
 CC useful in therapeutic, diagnostic and pharmacogenomic applications.
 CC NHP polynucleotides are used as hybridisation probes for screening
 CC libraries and assessing gene expression patterns. They can also be
 CC used for treating related biological disorders such as obesity, high
 CC blood pressure, arthritis, connective tissue disorders and infertility.
 CC They are also used in gene therapy.
 XX SQ Sequence 966 BP; 161 A; 316 C; 305 G; 184 T; 0 other;
 Query Match 31.7%; Score 903; DB 24; Length 966;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 57 TGAGCCAGAGCGGAGTAGTCTCCATCCGACTGGACCGGACATTAACGGCGCGG 116
 DB 57 TGAGCCAGAGCGGAGTAGTCTCCATCCGACTGGACCGGACATTAACGGCGCGG 116
 QY 117 CTACTACTGGGGGTCGCCAGGACTCCGGGGATCAGGACTATTTTCAGATCACAGC 176
 DB 117 CTACTACTGGGGGTCGCCAGGACTCCGGGGATCAGGACTATTTTCAGATCACAGC 176
 QY 177 ATTTTCAGGAGACTTTTACCTACACTGACCGCGGATGCTCAGTTCTTGCTCCCGGCTT 236
 DB 177 ATTTTCAGGAGACTTTTACCTACACTGACCGCGGATGCTCAGTTCTTGCTCCCGGCTT 236
 QY 237 CTCACCTAGCATCTGGGCTGCCCTTCCAGGGGCTCACGGGGGCTCTTCAGACCTGGG 296
 DB 237 CTCACCTAGCATCTGGGCTGCCCTTCCAGGGGCTCACGGGGGCTCTTCAGACCTGGG 296
 QY 297 ACCTGCTCTTATTTCTGGGACGTGAACCGCGGACCTGCTGCTGCTGTAGCCT 356
 DB 297 ACCTGCTCTTATTTCTGGGACGTGAACCGCGGACCTGCTGCTGCTGTAGCCT 356
 QY 357 GTGGGGGGGCTCCGGAGCGCTTTGGCTACCGAGGCGCGAGTAGTATGATTCATTCAGCCGCT 416
 DB 357 GTGGGGGGGCTCCGGAGCGCTTTGGCTACCGAGGCGCGAGTAGTATGATTCATTCAGCCGCT 416
 QY 417 GCGCAATGCTAGCGCGCGGGGAGCGGACACAGCGAGGGGCGACACCTTCTTCAGCGG 476
 DB 417 GCGCAATGCTAGCGCGCGGGGAGCGGACACAGCGAGGGGCGACACCTTCTTCAGCGG 476
 QY 477 CCGGGGTGTTCCGGGGGGGCTTCCGGAGAGCCACCTCTGCTGCGGGGTGCGGCTCGG 536

DB 477 CCGGGGTGTTCCGGGGGGGCTTCCGGAGAGCCACCTCTGCTGCGGGGTGGCCTCGGG 536
 QY 537 CTGGAACCCCGCCATCTACGGGCGCTGACCCCTTACAGCGCGCGGGGGCTTCGG 596
 DB 537 CTGGAACCCCGCCATCTACGGGCGCTGACCCCTTACAGCGCGCGGGGGCTTCGG 596
 QY 597 GGAGAGTCGTAGCCGCGCAGGCTCTGGGCGCGCAAGCGTTTCTGTTCTTCTCC 656
 DB 597 GGAGAGTCGTAGCCGCGCAGGCTCTGGGCGCGCAAGCGTTTCTGTTCTTCTCC 656
 QY 657 CGTGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 716
 DB 657 CGTGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 716
 QY 717 ACATTATCTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 776
 DB 717 ACATTATCTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 776
 QY 777 CAACCCCATCAACATCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 836
 DB 777 CAACCCCATCAACATCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 836
 QY 837 CAAGGTCAACGGCAATGCGGCGCTGACGCTGCGCAACTTCTGCTGCTGCGACAAGCT 896
 DB 837 CAAGGTCAACGGCAATGCGGCGCTGACGCTGCGCAACTTCTGCTGCTGCGACAAGCT 896
 QY 897 GAACAAAGTCACTGACAGCAAGCCCGGAGTACTGGGACACTGCCATCTTCCACGAGCA 956
 DB 897 GAACAAAGTCACTGACAGCAAGCCCGGAGTACTGGGACACTGCCATCTTCCACGAGCA 956
 QY 957 GGA 959
 DB 957 GGA 959
 RESULT 8
 ID AAA95831 standard; cDNA; 1143 BP.
 XX AC AAA95831;
 XX DT 23-FEB-2001 (first entry)
 XX DE Human metalloproteinase ADAMTS-5 cDNA.
 XX KW Human; ADAMTS-5; metalloproteinase; ADAM;
 KW KW a disintegrin and metalloproteinase domain; thrombospondin domain;
 KW KW vaccine; neutropic; neuroprotective; antiparkinsonian;
 KW KW cerebroprotective; cyostatic; antiarthritic; immunosuppressive;
 KW KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
 KW KW autoimmune disease; brain tumour; brain injury; ss.
 XX OS Homo sapiens.
 XX PH Key Location/Qualifiers
 XX FT CDS 1..1143
 XX FT /*tag= a
 XX FT /partial
 XX FT /product= "ADAMTS-5"
 XX PN WO200053774-A2.
 XX PF 14-SEP-2000.
 XX PR 08-MAR-2000; 2000WO-US06237.
 XX PR 08-MAR-1999; 99US-0264585.
 XX PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
 XX PI Kelner GS, Clark M, Makl RA;
 XX

KW osteopathic; antiarthritic; primer; adaptor; ss.

OS Homo sapiens.

PN WO200134785-A1.

XX 17-MAY-2001.

XX 10-NOV-2000; 2000WO-JP07917.

XX 11-NOV-1999; 99JP-0321740.

PR 16-MAY-2000; 2000JP-0144020.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

PA (KAZU-) KAZUSA DNA RES INST.

XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

XX Metalloprotease with aggrecanase activity for treating joint diseases especially osteoarthritis -

XX Example 13; Page 66; 85pp; Japanese.

CC This invention relates to a metalloprotease with aggrecanase activity. The invention includes protein and DNA sequences of the metalloprotease, vectors containing the DNA, host cells transformed by the vectors, and antibodies directed against the metalloprotease. The antibodies, protein and DNA sequences can be used in the treatment and prevention of joint diseases, particularly osteoarthritis. The treatment may result in osteopathic and antiarthritic activity. The present sequence represents an adaptor primer used in the isolation and characterisation of the metalloprotease gene of the invention.

XX Sequence 30 BP; 8 A; 6 C; 10 G; 6 T; 0 other;

Query Match 1.1%; Score 30; DB 22; Length 30;

Best Local Similarity 100.0%; Pred. NO. 0.0017;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 TTTACCTACACCTGACCGCGGATGCTCAGT 220

|||||

Db 30 TTTACCTACACCTGACCGCGGATGCTCAGT 1

RESULT 11

AAH41030/C

ID AAH41030 standard; DNA; 30 BP.

XX AC AAH41030;

XX 23-AUG-2001 (first entry)

DE Adaptor primer SEQ ID 23 used in metalloprotease DNA isolation.

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

KW osteopathic; antiarthritic; primer; adaptor; ss.

OS Homo sapiens.

XX WO200134785-A1.

XX 17-MAY-2001.

XX 10-NOV-2000; 2000WO-JP07917.

XX 11-NOV-1999; 99JP-0321740.

PR 16-MAY-2000; 2000JP-0144020.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

PA (KAZU-) KAZUSA DNA RES INST.

XX

PI Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

XX Metalloprotease with aggrecanase activity for treating joint diseases especially osteoarthritis -

XX Example 13; Page 66; 85pp; Japanese.

XX This invention relates to a metalloprotease with aggrecanase activity. The invention includes protein and DNA sequences of the metalloprotease, vectors containing the DNA, host cells transformed by the vectors, and antibodies directed against the metalloprotease. The antibodies, protein and DNA sequences can be used in the treatment and prevention of joint diseases, particularly osteoarthritis. The treatment may result in osteopathic and antiarthritic activity. The present sequence represents an adaptor primer used in the isolation and characterisation of the metalloprotease gene of the invention.

XX Sequence 30 BP; 10 A; 6 C; 6 G; 8 T; 0 other;

Query Match 1.1%; Score 30; DB 22; Length 30;

Best Local Similarity 100.0%; Pred. NO. 0.0017;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 TTTTTCAGATCACAGCATTTTCAGGAGGACT 190

|||||

Db 30 TTTTTCAGATCACAGCATTTTCAGGAGGACT 1

RESULT 12

AAH41017/C

ID AAH41017 standard; DNA; 41 BP.

XX AC AAH41017;

XX 23-AUG-2001 (first entry)

DE PCR primer specific for human metalloprotease DNA.

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

KW osteopathic; antiarthritic; PCR primer; ss.

OS Homo sapiens.

PN WO200134785-A1.

XX 17-MAY-2001.

XX 10-NOV-2000; 2000WO-JP07917.

XX 11-NOV-1999; 99JP-0321740.

PR 16-MAY-2000; 2000JP-0144020.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

PA (KAZU-) KAZUSA DNA RES INST.

XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

XX Metalloprotease with aggrecanase activity for treating joint diseases especially osteoarthritis -

XX Example 4; Page 62; 85pp; Japanese.

XX This invention relates to a metalloprotease with aggrecanase activity. The invention includes protein and DNA sequences of the metalloprotease, vectors containing the DNA, host cells transformed by the vectors, and antibodies directed against the metalloprotease. The antibodies, protein and DNA sequences can be used in the treatment and prevention of joint diseases, particularly osteoarthritis. The treatment may result in osteopathic and antiarthritic activity. The present sequence represents

CC a PCR primer used in the isolation and characterisation of the
XX metalloprotease gene of the invention.

SQ Sequence 41 BP; 5 A; 16 C; 13 G; 7 T; 0 other;

Query Match 1.1%; Score 30; DB 22; Length 41;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1720 GCCTCCGGAAGAGCTTCGGGAGGAGCAG 1749

Db 41 GCCTCCGGAAGAGCTTCGGGAGGAGCAG 12

RESULT 13

AAH41016
ID AAH41016 standard; DNA; 42 BP.

XX AC AAH41016;

DT 23-AUG-2001 (first entry)

DE PCR primer specific for human metalloprotease DNA.

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

KW osteopathic; antiarthritic; PCR primer; ss.

XX OS Homo sapiens.

PN WO200134785-A1.

PD 17-MAY-2001.

PF 10-NOV-2000; 2000WO-JP07917.

PR 11-NOV-1999; 99JP-0321740.

PR 16-MAY-2000; 2000JP-0144020.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

PA (KAZU-) KAZUSA DNA RES INST.

XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

XX Metalloprotease with aggrecanase activity for treating joint diseases
PT especially osteoarthritis

PS Example 4; Page 62; 85pp; Japanese.

CC This invention relates to a metalloprotease with aggrecanase activity.
CC The invention includes protein and DNA sequences of the metalloprotease.
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC a PCR primer used in the isolation and characterisation of the
CC metalloprotease gene of the invention.

SQ Sequence 42 BP; 6 A; 12 C; 11 G; 13 T; 0 other;

Query Match 1.1%; Score 30; DB 22; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTTTTCTGGGCATCTACCTGGCT 30

Db 13 ATGCTTTTCTGGGCATCTACCTGGCT 42

RESULT 14

AAH41018

ID AAH41018 standard; DNA; 27 BP.

XX AC AAH41018;

DT 23-AUG-2001 (first entry)

DE PCR primer specific for human metalloprotease DNA SEQ ID 9.

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

KW osteopathic; antiarthritic; PCR primer; ss.

XX OS Homo sapiens.

PN WO200134785-A1.

PD 17-MAY-2001.

PF 10-NOV-2000; 2000WO-JP07917.

PR 11-NOV-1999; 99JP-0321740.

PR 16-MAY-2000; 2000JP-0144020.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

PA (KAZU-) KAZUSA DNA RES INST.

XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

XX Metalloprotease with aggrecanase activity for treating joint diseases
PT especially osteoarthritis

PS Example 5; Page 63; 85pp; Japanese.

CC This invention relates to a metalloprotease with aggrecanase activity.
CC The invention includes protein and DNA sequences of the metalloprotease.
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC a PCR primer used in the isolation and characterisation of the
CC metalloprotease gene of the invention.

SQ Sequence 27 BP; 5 A; 6 C; 11 G; 5 T; 0 other;

Query Match 0.9%; Score 27; DB 22; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1534 AAGCACAGGGTGGATGCTCTGGGCC 1560

Db 1 AAGCACAGGGTGGATGCTCTGGGCC 27

RESULT 15

AAH41019/c

ID AAH41019 standard; DNA; 37 BP.

XX AC AAH41019;

DT 23-AUG-2001 (first entry)

DE PCR primer specific for human metalloprotease DNA SEQ ID 10.

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

KW osteopathic; antiarthritic; PCR primer; ss.

XX OS Homo sapiens.

PN WO200134785-A1.

PD 17-MAY-2001.

```

XX 10-NOV-2000; 2000WO-JP07917.
XX
XX 11-NOV-1999; 99JP-0321740.
XX
XX 16-MAY-2000; 2000JP-0144020.
XX
XX (YAMA ) YAMANOUCHI PHARM CO LTD.
XX (KAZU-) KAZUSA DNA RES INST.
XX
XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX
XX WPI; 2001-343602/36.
XX
XX Metalloprotease with aggreganase activity for treating joint diseases
XX especially osteoarthritis
XX
XX Example 5; Page 63; 85pp; Japanese.
XX
XX This invention relates to a metalloprotease with aggreganase activity.
XX The invention includes protein and DNA sequences of the metalloprotease,
XX vectors containing the DNA, host cells transformed by the vectors, and
XX antibodies directed against the metalloprotease. The antibodies, protein
XX and DNA sequences can be used in the treatment and prevention of joint
XX diseases, particularly osteoarthritis. The treatment may result in
XX osteopathic and antiarthritic activity. The present sequence represents
XX a PCR primer used in the isolation and characterisation of the
XX metalloprotease gene of the invention.
XX
XX Sequence 37 BP; 7 A; 14 C; 14 G; 2 T; 0 other;
XX
XX Query Match 0.9%; Score 27; DB 22; Length 37;
XX Best Local Similarity 100.0%; Pred. No. 0.042;
XX Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 2824 CTGGACTTCTGCGTCTGAGGCGGTGC 2850
XX |||||||||||||||||||||||
XX Db 37 CTGGACTTCTGCGTCTGAGGCGGTGC 11
XX
XX Search completed: May 2, 2003, 01:10:45
XX Job time : 622 secs

```

1. The first part of the paper is devoted to the study of the properties of the function $f(x)$ defined by the equation

$$f(x) = \int_0^x \frac{1}{1+t^2} dt$$

for $x \in \mathbb{R}$. It is shown that $f(x)$ is an odd function and that $f(x) \in C^1(\mathbb{R})$. Moreover, it is proved that $f(x)$ is a strictly increasing function and that $f(x) \in C^2(\mathbb{R})$. Finally, it is shown that $f(x)$ is a concave function.

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 22:52:52 ; Search time 103 Seconds
(without alignments)
8494.649 Million cell updates/sec

Title: US-10-009-332-2

Perfect score: 2853

Sequence: 1 atgtcttctgctggcctctct.....gcgtctgagcgctgctga 2853

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- Issued_Patents_NA:*
- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
 - 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
 - 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
 - 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
 - 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
 - 6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	0.8	23	2114	4	US-09-130-491-7
2	0.7	19	4533	3	US-08-726-214-5
3	0.6	18	311	4	US-09-643-597-314
4	0.6	18	2457	4	US-08-872-757-1
5	0.6	18	2487	1	US-08-377-232-1
6	0.6	18	2830	2	US-09-010-928B-1
7	0.6	18	3226	1	US-07-862-021B-11
8	0.6	18	3226	1	US-08-313-288B-11
9	0.6	18	3226	5	PCT-US93-03164-11
10	0.6	18	3546	4	US-08-872-757-3
11	0.6	18	4149	2	US-08-737-715-1
12	0.6	18	4403765	4	US-09-103-840A-2
13	0.6	18	4403765	4	US-09-103-840A-2
14	0.6	18	4411529	4	US-09-103-840A-1
15	0.6	18	4411529	4	US-09-103-840A-1
16	0.6	17	426	4	US-08-914-375C-55
17	0.6	17	438	1	US-08-540-242A-5
18	0.6	17	438	5	PCT-US96-01471-5
19	0.6	17	444	1	US-08-688-609-7
20	0.6	17	444	3	US-09-002-832-7
21	0.6	17	445	4	US-08-688-908-6
22	0.6	17	504	1	US-08-540-242A-1
23	0.6	17	504	4	US-08-692-922-6
24	0.6	17	504	5	PCT-US96-01471-1
25	0.6	17	701	3	US-08-292-345B-1
26	0.6	17	702	2	US-09-003-081-1
27	0.6	17	702	3	US-08-648-262-1

28	17	0.6	702	3	US-08-648-263-1	Sequence 1, Appli
29	17	0.6	900	2	US-08-518-862C-5	Sequence 5, Appli
30	17	0.6	1205	3	US-09-120-772-1	Sequence 1, Appli
31	17	0.6	1452	1	US-08-187-785-2	Sequence 2, Appli
32	17	0.6	1480	4	US-09-142-569-1	Sequence 1, Appli
33	17	0.6	1659	1	US-08-333-358-7	Sequence 7, Appli
34	17	0.6	1659	1	US-08-463-694-7	Sequence 7, Appli
35	17	0.6	1659	1	US-08-694-501-7	Sequence 7, Appli
36	17	0.6	2023	4	US-09-288-143-54	Sequence 54, Appli
37	17	0.6	2793	2	US-08-347-563A-1	Sequence 1, Appli
38	17	0.6	2793	3	US-08-485-942A-1	Sequence 1, Appli
39	17	0.6	2793	3	US-08-488-214A-1	Sequence 1, Appli
40	17	0.6	2793	3	US-08-488-308A-1	Sequence 1, Appli
41	17	0.6	2793	4	US-08-483-211A-1	Sequence 1, Appli
42	17	0.6	2793	4	US-08-488-223A-1	Sequence 1, Appli
43	17	0.6	2793	4	US-08-438-431A-1	Sequence 1, Appli
44	17	0.6	2855	2	US-08-852-153-1	Sequence 1, Appli
45	17	0.6	3255	2	US-08-852-153-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-130-491-7
; Sequence 7, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:

; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-75, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2114
; TYPE: DNA
; ORGANISM: Rattus rattus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1445)
US-09-130-491-7

Query Match 0.8%; Score 23; DB 4; Length 2114;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2008 GACAACTGTGGGGTGTGGGG 2030
|||||
Db 807 GACAACTGTGGGGTGTGGGG 829

RESULT 2
US-08-726-214-5
; Sequence 5, Application US/08726214
; Patent No. 6107076
; GENERAL INFORMATION:

; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gilman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas

COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,214
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/005,498
FILING DATE: 04-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: UTSD:450
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4533 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-726-214-5

Query Match 0.7%; Score 19; DB 3; Length 4533;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGGGCTCCAAGAAGAGA 2004
|||||
DB 3078 CCTGGGCTCCAAGAAGAGA 3096

RESULT 3
US-09-643-597-314/c
Sequence 314, Application US/09643597
Patent No. 6426072
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Lijun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.455C11
CURRENT APPLICATION NUMBER: US/09/643,597
CURRENT FILING DATE: 2000-08-21
NUMBER OF SEQ ID NOS: 369
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 314
LENGTH: 311
TYPE: DNA
ORGANISM: Homo sapien
US-09-643-597-314

Query Match 0.6%; Score 18; DB 4; Length 311;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1988 TGGGCTCCAAGAAGAGAT 2005
|||||

Db 86 TGGGCTCCAAGAAGAGAT 69
RESULT 4
US-08-872-757-1
Sequence 1, Application US/08872757
Patent No. 6258584
GENERAL INFORMATION:
APPLICANT: Prockop, Darwin J.
APPLICANT: Hojima, Yoshio
APPLICANT: Li, Shi-Wu
APPLICANT: Sieron, Aleksander
TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
METHODS OF SEQUENCES: 6
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/872,757
FILING DATE: 10-JUN-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/609,187
FILING DATE: 01-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8389-028-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2457 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2190
US-08-872-757-1

Query Match 0.6%; Score 18; DB 4; Length 2457;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1080 CCACGAGCTGGGCCACGT 1097
|||||

Db 636 CCACGAGCTGGGCCACGT 653

RESULT 5
US-08-377-292-1
Sequence 1, Application US/08377292
Patent No. 5693615
GENERAL INFORMATION:
APPLICANT: STONE, ROGER L.
TITLE OF INVENTION: THERAPEUTIC FORMULAS FOR OSTEOINDUCTION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Procter & Gamble Company

```

; STREET: 11810 East Miami River Road
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 45239-8707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/377,292
; FILING DATE: 23-JAN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/243,435
; FILING DATE:
; APPLICATION NUMBER: US/08/117,367
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Corstanje, Brahm J.
; REGISTRATION NUMBER: 34,804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 513-245-2858
; TELEFAX: 513-741-3012
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2487 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-377-292-1

```

```

Query Match 0.6%; Score 18; DB 1; Length 2487;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1080 CCAGGAGCTGGCCACGT 1097
Db 665 CCAGGAGCTGGCCACGT 682

```

```

RESULT 6
US-09-010-928B-1
; Sequence 1, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Havashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P

```

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2830
; OTHER INFORMATION: /note= "Flagelliform DNA sequence
; OTHER INFORMATION: taken from the 5' region. The putative start codon is at
; OTHER INFORMATION: position 219"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 219..2830
; US-09-010-928B-1

```

```

Query Match 0.6%; Score 18; DB 2; Length 2830;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2234 CGGTGAGCGGGACCTGG 2251
Db 2384 CGGTGAGCGGGACCTGG 2401

```

```

RESULT 7
US-07-862-021B-11
; Sequence 11, Application US/07862021B
; Patent No. 5279966
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,021B
; FILING DATE: 19920405
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543
; US-07-862-021B-11

```

Query Match 0.6%; Score 18; DB 1; Length 3226;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1203 CCCCTGGTCAGCCTGCAG 1220
|||||
Db 1467 CCCCTGGTCAGCCTGCAG 1484

RESULT 8

US-08-313-288B-11
; Sequence 11, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:

INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543

US-08-313-288B-11

Query Match 0.6%; Score 18; DB 1; Length 3226;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1203 CCCCTGGTCAGCCTGCAG 1220
|||||
Db 1467 CCCCTGGTCAGCCTGCAG 1484

RESULT 9

PCT-US93-03164-11
; Sequence 11, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543
PCT-US93-03164-11

Query Match 0.6%; Score 18; DB 5; Length 3226;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1203 CCCCTGGTCAGCCTGCAG 1220
|||||
Db 1467 CCCCTGGTCAGCCTGCAG 1484

RESULT 10

US-08-872-757-3
; Sequence 3, Application US/08872757
; Patent No. 6258584
; GENERAL INFORMATION:
; APPLICANT: Prockop, Darwin J.
; APPLICANT: Hojima, Yoshio
; APPLICANT: Li, Shi-Wu
; APPLICANT: Sieron, Aleksander
; TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
; TITLE OF INVENTION: PROCESSES, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/872,757
FILING DATE: 10-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/609,187
FILING DATE: 01-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8389-028-999
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3546 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2958
US-08-872-757-3

Query Match 0.6%; Score 18; DB 4; Length 3546;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1080 CCACGAGCTGGCCACGT 1097
|||||

Db 636 CCACGAGCTGGCCACGT 653

RESULT 11

US-08-737-715-1/c
Sequence 1, Application US/08/737,715
Patent No. 5958685
GENERAL INFORMATION:
APPLICANT: Ebina, Yousuke
TITLE OF INVENTION: MUTANT HUMAN INSULIN RECEPTOR DNA
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,715
FILING DATE: 12-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 134827/1995
FILING DATE: 12-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Nakamura, Dean H.
REGISTRATION NUMBER: 33,981
REFERENCE/DOCKET NUMBER: Q-43323
TELEPHONE: (202)293-7060
TELEFAX: (202)293-7860
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4149 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4146
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 82..4146
US-08-737-715-1

Query Match 0.6%; Score 18; DB 2; Length 4149;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1701 GGAGCCCTGCCCGAGCTC 1718
|||||

Db 3099 GGAGCCCTGCCCGAGCTC 3082

RESULT 12

US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 0.6%; Score 18; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 433 CCGGCGGCGCAGCGCAAC 450
|||||

Db 77489 CCGGCGGCGCAGCGCAAC 77506

RESULT 13

US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 0.6%; Score 18; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2225 TGGTGTGCGGGTGGAGC 2242
|||||
Db 3375303 TGGTGTGCGGGTGGAGC 3375286

RESULT 14
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 0.6%; Score 18; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 433 CCGGCGGCGGCGCAAC 450
|||||
Db 77510 CCGGCGGCGGCGCAAC 77527

RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 0.6%; Score 18; DB 4; Length 4411529;

Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2225 TGGTGTGCGGGTGGAGC 2242
|||||
Db 3379611 TGGTGTGCGGGTGGAGC 3379594

Search completed: May 2, 2003, 05:13:52
Job time : 14589 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic : nucleic search, using sw model

Run on: May 2, 2003, 00:35:12 ; Search time 339 Seconds
(without alignments)
9944.440 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgcttttgcggcatcct.....gcgtcctgagccgtgcta 2853

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 746064 seqs, 590810554 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2695	94.5	2853	10 US-09-965-631-3	Sequence 3, Appli
2	2695	94.5	3446	10 US-09-965-631-7	Sequence 7, Appli
3	1435	50.3	2469	9 US-10-163-316-3	Sequence 3, Appli
4	1435	50.3	2940	9 US-10-163-316-1	Sequence 1, Appli
5	1035	36.3	1104	10 US-09-965-631-5	Sequence 5, Appli
6	903	31.7	966	10 US-09-965-631-1	Sequence 1, Appli
7	23	0.8	2114	12 US-10-105-929-7	Sequence 1, Appli
c 8	22	0.8	1813	9 US-09-931-457A-10	Sequence 10, Appl
c 9	21	0.7	506	9 US-09-918-995-22019	Sequence 22019, A
c 10	21	0.7	1827	10 US-09-925-297-180	Sequence 180, Appl
c 11	20	0.7	1017	9 US-09-938-842A-60	Sequence 60, Appl
c 12	19	0.7	256	9 US-09-535-459-1835	Sequence 1835, Ap
c 13	19	0.7	408	9 US-09-918-995-15962	Sequence 15962, A
c 14	19	0.7	431	10 US-09-925-297-85	Sequence 85, Appl
c 15	19	0.7	452	9 US-09-918-995-32286	Sequence 32286, A
c 16	19	0.7	466	9 US-09-918-995-34926	Sequence 34926, A
c 17	19	0.7	469	9 US-09-918-995-11410	Sequence 11410, A
c 18	18	0.6	157	10 US-09-864-761-19861	Sequence 19861, A
c 19	18	0.6	258	10 US-09-864-761-29711	Sequence 29711, A

c 20	18	0.6	311	10 US-09-735-705-314	Sequence 314, App
c 21	18	0.6	311	10 US-09-850-716A-314	Sequence 314, App
c 22	18	0.6	311	10 US-09-897-778-314	Sequence 314, App
c 23	18	0.6	381	10 US-09-878-574-669	Sequence 669, App
c 24	18	0.6	401	9 US-09-954-531-91	Sequence 91, Appli
c 25	18	0.6	401	10 US-09-954-456-752	Sequence 752, App
c 26	18	0.6	462	10 US-09-864-761-3080	Sequence 3080, Ap
c 27	18	0.6	471	9 US-09-918-995-27364	Sequence 27364, A
c 28	18	0.6	482	9 US-10-040-739-149	Sequence 149, App
c 29	18	0.6	490	9 US-09-918-995-30714	Sequence 30714, A
c 30	18	0.6	496	9 US-09-918-995-11026	Sequence 11026, A
c 31	18	0.6	527	9 US-09-918-995-31632	Sequence 31632, A
c 32	18	0.6	597	10 US-09-864-761-13150	Sequence 13150, A
c 33	18	0.6	682	10 US-09-833-381-836	Sequence 836, App
c 34	18	0.6	954	10 US-09-815-242-9757	Sequence 9757, Ap
c 35	18	0.6	1014	9 US-10-102-806-305	Sequence 305, App
c 36	18	0.6	1376	9 US-10-011-445-18	Sequence 18, Appli
c 37	18	0.6	1410	10 US-09-779-448-3	Sequence 3, Appli
c 38	18	0.6	1536	10 US-09-779-448-4	Sequence 4, Appli
c 39	18	0.6	1859	10 US-09-779-448-2	Sequence 2, Appli
c 40	18	0.6	1953	10 US-09-784-358-9	Sequence 9, Appli
c 41	18	0.6	2175	10 US-09-784-358-7	Sequence 7, Appli
c 42	18	0.6	2316	10 US-09-784-358-13	Sequence 13, Appli
c 43	18	0.6	2457	10 US-09-850-048A-1	Sequence 1, Appli
c 44	18	0.6	2489	10 US-09-764-864-552	Sequence 552, App
c 45	18	0.6	2504	10 US-09-764-864-103	Sequence 103, App

ALIGNMENTS

RESULT 1

US-09-965-631-3

Sequence 3, Application US/09965631

Patent No. US20020115842A1

GENERAL INFORMATION:

APPLICANT: Friddle, Carl Johan

APPLICANT: Hilbun, Erin

TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Enco

FILE REFERENCE: LEX-0241-USA

CURRENT APPLICATION NUMBER: US/09/965,631

CURRENT FILING DATE: 2001-09-27

PRIOR FILING DATE: 2000-09-29

NUMBER OF SEQ ID NOS: 7

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

TYPE: DNA

ORGANISM: homo sapiens

US-09-965-631-3

Query Match	94.5%	Score 2695;	DB 10;	Length 2853;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 2795;	Conservative	0;	Mismatches	2;
			Indels	0;
			Gaps	0;
Qy	57	TGAGCCAGAGCGGAGGTACTGTTCCATCCGACGTGGACCGGACATTAAACGGCGCG 116		
Db	57	TGAGCCAGAGCGGAGGTACTGTTCCATCCGACGTGGACCGGACATTAAACGGCGCG 116		
Qy	117	CTACTACTGCGGGGTCCCGAGGACTCCGGGGATCAGGGACTCATTTTTCAGATCAGC 176		
Db	117	CTACTACTGCGGGGTCCCGAGGACTCCGGGGATCAGGGACTCATTTTTCAGATCAGC 176		
Qy	177	ATTTCAGGAGACTTTTACCTACCTGAGCGCGGATGCTCAGTCTTGTGCTCCCGCCT 236		
Db	177	ATTTCAGGAGACTTTTACCTACCTGAGCGCGGATGCTCAGTCTTGTGCTCCCGCCT 236		
Qy	237	CTCCACTGAGCATCTGGCGCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGG 296		
Db	237	CTCCACTGAGCATCTGGCGCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGG 296		
Qy	297	ACGCTGCTTCTATTCTCGGACGCTGAACGCCGAGCGGAGCTGCTGCTGCTGCTG 356		

|||||
Db 297 ACCTGCTTCTATTTCTGGGACGCTGAACGCCGAGCCGACTCGTTTCGCTGCTGAGCCCT 356
QY 357 GTCCGGGGGCTCCGGGAGCCCTTTGGCTACCGAGGCGCGGAGTATGTCTATTAGCCCGCT 416
Db 357 GTCCGGGGGCTCCGGGAGCCCTTTGGCTACCGAGGCGCGGAGTATGTCTATTAGCCCGCT 416
QY 417 GCCCAATGCTAGCGCGCGCGCGGCGAGCGACACACGCGAGGGGCGACACTTCTCCAGCG 476
Db 417 GCCCAATGCTAGCGCGCGCGCGGCGAGCGACACACGCGAGGGGCGACACTTCTCCAGCG 476
QY 477 CCGGGGTGTTCCCGGGGGGCTTCCGGAGACCCACCTCTCGCTCGGGGTGGCTCGGG 536
Db 477 CCGGGGTGTTCCCGGGGGGCTTCCGGAGACCCACCTCTCGCTCGGGGTGGCTCGGG 536
QY 537 CTGGAACCCCGCCATCCTACGGGCCCTTGAACCCCTTACAAGCCGCGCGGGGCTTCGG 596
Db 537 CTGGAACCCCGCCATCCTACGGGCCCTTGAACCCCTTACAAGCCGCGCGGGGCTTCGG 596
QY 597 GGAGAGTCGTAGCCGGGCGAGTCTGGGCGCGCCAGCGTTTCGTGTCTATCCCGGGTA 656
Db 597 GGAGAGTCGTAGCCGGGCGAGTCTGGGCGCGCCAGCGTTTCGTGTCTATCCCGGGTA 656
QY 657 CGTGGAGACGCTGGTGTGCGGACGAGTCAATGCTCAAGTTCCACGGCGCGGACCTGGA 716
Db 657 CGTGGAGACGCTGGTGTGCGGACGAGTCAATGCTCAAGTTCCACGGCGCGGACCTGGA 716
QY 717 ACATTATCTGCTGACGCTGCTGCAACGCGGCGGCGGACTCTACCGGCATCCAGCATCCT 776
Db 717 ACATTATCTGCTGACGCTGCTGCAACGCGGCGGCGGACTCTACCGGCATCCAGCATCCT 776
QY 777 CAACCCCATCAACATGTTGTGTCAAGTGTCTTTCAGATCTGATCTCCGGG 836
Db 777 CAACCCCATCAACATGTTGTGTCAAGTGTCTTTCAGATCTGATCTCCGGG 836
QY 837 CAAGGTCACCGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGCGCAGAAGCT 896
Db 837 CAAGGTCACCGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGCGCAGAAGCT 896
QY 897 GAACAAAGTGTGACAAAGACCCCGAGTACTGGGACACTGCCATCTCTTACCAGGCA 956
Db 897 GAACAAAGTGTGACAAAGACCCCGAGTACTGGGACACTGCCATCTCTTACCAGGCA 956
QY 957 GGACCTGTGTGGAGCCACCACTGTACACCTGGGCGGCTGATGGGTACCATGTG 1016
Db 957 GGACCTGTGTGGAGCCACCACTGTACACCTGGGCGGCTGATGGGTACCATGTG 1016
QY 1017 TGACCCCAAGAGAAAGTGTCTGTCAATTGAGGACGATGGGCTTCCATCAGCCCTTACCAC 1076
Db 1017 TGACCCCAAGAGAAAGTGTCTGTCAATTGAGGACGATGGGCTTCCATCAGCCCTTACCAC 1076
QY 1077 TGCCCAAGAGCTGGGCCAGGTGTTTCAACATGCCCCATGACAATGTGAAGTCTGTGAGGA 1136
Db 1077 TGCCCAAGAGCTGGGCCAGGTGTTTCAACATGCCCCATGACAATGTGAAGTCTGTGAGGA 1136
QY 1137 GGTGTTTGGGAAGCTCCGAGCCCAACACATGATGTCCCGACCCCTATCCAGATCGACCG 1196
Db 1137 GGTGTTTGGGAAGCTCCGAGCCCAACACATGATGTCCCGACCCCTATCCAGATCGACCG 1196
QY 1197 TGCCAAACCCCTGTGACGCTGCAAGTGTGCCATCATCACCAGTCTTCCGAGCGGGCA 1256
Db 1197 TGCCAAACCCCTGTGACGCTGCAAGTGTGCCATCATCACCAGTCTTCCGAGCGGGCA 1256
QY 1257 CGGTGACTGCTCTGTGGACCAACCAAGCCCATCTCCCTGCGCCGAGGATCTCCCGGG 1316
Db 1257 CGGTGACTGCTCTGTGGACCAACCAAGCCCATCTCCCTGCGCCGAGGATCTCCCGGG 1316
QY 1317 CGCCAGCTACACCTGTAGCCAGCAGTGGAGCTGTGCTTGTGGCTGGGCTCCAGCCCTG 1376
Db 1317 CGCCAGCTACACCTGTAGCCAGCAGTGGAGCTGTGCTTGTGGCTGGGCTCCAGCCCTG 1376
QY 1377 TCTTACATGCAAGTACTGCACCAAGCTGTGTGACCCGGGAGGCCAAGGACAGATGGT 1436
|||||

Db 1377 TCCTTACATGCAAGTACTGCACCAAGCTGTGTGCAACCGGGAAGGCCAAGGACAGATGGT 1436
QY 1437 GTCCGAGACCCGCCACTTCCCTGGGCGCATGGCACCAAGCTGTGGCGAGGCAAGCTCTG 1496
Db 1437 GTCCGAGACCCGCCACTTCCCTGGGCGCATGGCACCAAGCTGTGGCGAGGCAAGCTCTG 1496
QY 1497 CCTCAAGGGGCTCGTGGAGAGACACAACCTCAACAGCACAGGCTGAGTGTCTCTG 1556
Db 1497 CCTCAAGGGGCTCGTGGAGAGACACAACCTCAACAGCACAGGCTGAGTGTCTCTG 1556
QY 1557 GSCCAATGGATCCCTATGCCCCTGCTCGGCACATGTGTGGGCGCTGCAAGCTGGC 1616
Db 1557 GSCCAATGGATCCCTATGCCCCTGCTCGGCACATGTGTGGGCGCTGCAAGCTGGC 1616
QY 1617 CAGGAGGAGTGCACCAACCCACCCCTGSCCAAGGGGGCAAGTACTGCCAGGAGTGA 1676
Db 1617 CAGGAGGAGTGCACCAACCCACCCCTGSCCAAGGGGGCAAGTACTGCCAGGAGTGA 1676
QY 1677 GGTGAATACCGATCTCTCAACCTGGAGCCCTGCCCAAGCTCAGCCTCCGGAAGAGT 1736
Db 1677 GGTGAATACCGATCTCTCAACCTGGAGCCCTGCCCAAGCTCAGCCTCCGGAAGAGT 1736
QY 1737 CCGGAGGAGCAGTGTGAGGCTTTCACGGCTACAACACAGCACCAACCGGCTCACTCT 1796
Db 1737 CCGGAGGAGCAGTGTGAGGCTTTCACGGCTACAACACAGCACCAACCGGCTCACTCT 1796
QY 1797 CGCCGTGCAATGGGTGCCCAAGTACTCCGGGCTGTCTCCCGGGACAAAGTCAAGCTCAT 1856
Db 1797 CGCCGTGCAATGGGTGCCCAAGTACTCCGGGCTGTCTCCCGGGACAAAGTCAAGCTCAT 1856
QY 1857 CTGCGGAGCCAAATGGCACTGGCTACTTCTATGTCTGGCACCCAAAGTGTGGTGGAC 1916
Db 1857 CTGCGGAGCCAAATGGCACTGGCTACTTCTATGTCTGGCACCCAAAGTGTGGTGGAC 1916
QY 1917 GCTGTGCTCTCTGACTCCACTCCGCTGTGTCCAAAGGCAAGTGCATCAAGCTGCTG 1976
Db 1917 GCTGTGCTCTCTGACTCCACTCCGCTGTGTCCAAAGGCAAGTGCATCAAGCTGCTG 1976
QY 1977 TGATGGGAACCTGGGCTCCAAAGAGAGATTGACAAGTGTGGGCTGTGTGGGGAGACAA 2036
Db 1977 TGATGGGAACCTGGGCTCCAAAGAGAGATTGACAAGTGTGGGCTGTGTGGGGAGACAA 2036
QY 2037 TAAGAGCTGAAGAAGTGTGACTGTCTTCCACCAAGCCCATGCTGCTACATTTCTG 2096
Db 2037 TAAGAGCTGAAGAAGTGTGACTGTCTTCCACCAAGCCCATGCTGCTACATTTCTG 2096
QY 2097 GGTGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGAGCGGCTTACAAAGGCT 2156
Db 2097 GGTGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGAGCGGCTTACAAAGGCT 2156
QY 2157 GATCGGGATGACAACCTACCTGGCTCTGAAGAACAGCAAGCAAGTACCTGCTCAACGG 2216
Db 2157 GATCGGGATGACAACCTACCTGGCTCTGAAGAACAGCAAGCAAGTACCTGCTCAACGG 2216
QY 2217 GCATTTGCTGTGTGTCGGGCTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGCGGTA 2276
Db 2217 GCATTTGCTGTGTGTCGGGCTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGCGGTA 2276
QY 2277 CAGCGGACCGGACACAGCGGTGGAGAGCCTTCCGCGCCCATCTCTGCAAGCCGCT 2336
Db 2277 CAGCGGACCGGACACAGCGGTGGAGAGCCTTCCGCGCCCATCTCTGCAAGCCGCT 2336
QY 2337 GACCGTGTGAGTCTCTCTGCTGGGGAAGATGACACCCGCCCGGGTCCGCTACTCTCTTA 2396
Db 2337 GACCGTGTGAGTCTCTCTGCTGGGGAAGATGACACCCGCCCGGGTCCGCTACTCTCTTA 2396
QY 2397 TCTGCCAAAGAGCTCGGGAGGACAGTCTCTCAATCCCAAGAGCCCGGGGACCCCTC 2456
Db 2397 TCTGCCAAAGAGCTCGGGAGGACAGTCTCTCAATCCCAAGAGCCCGGGGACCCCTC 2456
QY 2457 TGTCTTGCACAAAGCTCTCTGAGCTCTTCCAAACAGGCTGGAGCAGCGGACACAGGCC 2516
Db 2457 TGTCTTGCACAAAGCTCTCTGAGCTCTTCCAAACAGGCTGGAGCAGCGGACACAGGCC 2516

```

QY 2517 CCTGACGCTGGTGGTGGAGCTGGGGCGGTGCTCCGCGAGCTGGCGAGTGGCCT 2576
|
|
|
Db 2517 CCTGACGCTGGTGGTGGAGCTGGGGCGGTGCTCCGCGAGCTGGCGAGTGGCCT 2576
|
|
|
QY 2577 GCAGAAAGCGGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2636
|
|
|
Db 2577 GCAGAAAGCGGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2636
|
|
|
QY 2637 TGACAGCCCATCGCCCGTGGAGACACAGCTGGCGGGAGCCCTGCCCCACCTGGGAGCT 2696
|
|
|
Db 2637 TGACAGCCCATCGCCCGTGGAGACACAGCTGGCGGGAGCCCTGCCCCACCTGGGAGCT 2696
|
|
|
QY 2697 CAGCGCTGGTCAACCTGCTCCAAAGAGCTGGCGCGGGGATTTTCAGAGCGCTCACTCAA 2756
|
|
|
Db 2697 CAGCGCTGGTCAACCTGCTCCAAAGAGCTGGCGCGGGGATTTTCAGAGCGCTCACTCAA 2756
|
|
|
QY 2757 GTGTGTGGCGACGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2816
|
|
|
Db 2757 GTGTGTGGCGCACGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2816
|
|
|
QY 2817 CCAGGAGCTGGAGCTTCTGCTGCTGAGGCGGTGCTGA 2853
|
|
|
Db 2817 CCAGGAGCTGGAGCTTCTGCTGCTGAGGCGGTGCTGA 2853
|
|
|

```

RESULT 2

```

US-09-965-631-7
; Sequence 7, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Encodir
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 3446
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-965-631-7

```

```

Query Match 94.5%; Score 2695; DB 10; Length 3446;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2795; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 57 TCAGCCAGAGCGGGAGGTAGTCTCCATCGGAGTGGACCCGGACATTAACGGCCGCG 116
|
|
|
Db 453 TGAGCCAGAGCGGGAGGTAGTCTCCATCGGAGTGGACCCGGACATTAACGGCCGCG 512
|
|
|
QY 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTCAGATCACAGC 176
|
|
|
Db 513 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTCAGATCACAGC 572
|
|
|
QY 177 ATTTGAGGAGGACTTTTACCTACACCTGACCCGGATGCTCAGTTCTTGGCTCCCGCCTT 236
|
|
|
Db 573 ATTTGAGGAGGACTTTTACCTACACCTGACCCGGATGCTCAGTTCTTGGCTCCCGCCTT 632
|
|
|
QY 237 CTTCACTGAGCATCTGGCGGTCCCGCTCCAGGGGCTACCGGGGGCTCTTCAGACCTTGG 296
|
|
|
Db 633 CTTCACTGAGCATCTGGCGGTCCCGCTCCAGGGGCTACCGGGGGCTCTTCAGACCTTGG 692
|
|
|
QY 297 AGCTGCTTCTATTCTGGGAGCTGAAGCGCGAGCGGAGCTTCGCTGCTGTGAGCCT 356
|
|
|
Db 693 AGCTGCTTCTATTCTGGGAGCTGAAGCGCGAGCGGAGCTTCGCTGCTGTGAGCCT 752
|
|
|
QY 357 GTGCGGGGGCTCCGCGAGCCTTTGGCTACCGGGGCGCGAGTATGTCATTAGCCGCT 416
|
|
|

```

```

Db 753 GTGCGGGGGGTCCGCGGAGGCTTTGGCTACCGAGGCGCGAGTATGTTCATTAGCCGCT 812
|
|
|
QY 417 GCCCAATCTAGCGCGCGCGCGCGAGCGCAACAGCGCAGGCGCGACACCTTTCTCCAGCG 476
|
|
|
Db 813 GCCCAATCTAGCGCGCGCGCGCGAGCGCAACAGCGCAGGCGCGACACCTTTCTCCAGCG 872
|
|
|
QY 477 CGGGGTGTTCCGGCGCGCGCGCTTCGGAGACCCACCTCTCGCTGCGGGGTGGCTCGG 536
|
|
|
Db 873 CGGGGTGTTCCGGCGCGCGCGCTTCGGAGACCCACCTCTCGCTGCGGGGTGGCTCGG 932
|
|
|
QY 537 CTGAACCCCGCCATCTCTACGGGCGCTTACAGCCCTTACAGCGCGCGGGTGGCTTCGG 596
|
|
|
Db 933 CTGAACCCCGCCATCTCTACGGGCGCTTACAGCCCTTACAGCGCGCGGGTGGCTTCGG 992
|
|
|
QY 597 GGAGAGTGTAGCGCGCGCGAGTCTGGCGCGCGCAAGCTTTCGTCTATCCCGCGGTA 656
|
|
|
Db 993 GGAGAGTGTAGCGCGCGCGAGTCTGGCGCGCGCAAGCTTTCGTCTATCCCGCGGTA 1052
|
|
|
QY 657 CGTGAGAGCTGTGTGTCGGGAGGAGTCAATGTCAGTTCAGGTCACGGCGCGGAGCTGGA 716
|
|
|
Db 1053 CGTGAGAGCTGTGTGTCGGGAGGAGTCAATGTCAGTTCAGGTCACGGCGCGGAGCTGGA 1112
|
|
|
QY 717 ACATTATCTGTGACGCTGTGGCAAGCGCGCGCGAGTCTTACCGCCATCTCAGTATCCT 776
|
|
|
Db 1113 ACATTATCTGTGACGCTGTGGCAAGCGCGCGCGAGTCTTACCGCCATCTCAGTATCCT 1172
|
|
|
QY 777 CAACCCCATCAACATCTGTGTGTCAGGTCGCTCTTTAGAGATCGTGA2TCGGGCG 836
|
|
|
Db 1173 CAACCCCATCAACATCTGTGTGTCAGGTCGCTCTTTAGAGATCGTGA2TCGGGCG 1232
|
|
|
QY 837 CAAGGTCAACCGCAATCGGCGCTGACGCTGCGCAACTCTGTGCTTGGCAGGAAGAGCT 896
|
|
|
Db 1233 CAAGGTCAACCGCAATCGGCGCTGACGCTGCGCAACTCTGTGCTTGGCAGGAAGAGCT 1292
|
|
|
QY 897 GAACAAAGTGTGAGTGAACAGCACCCGAGTACTGGGACACTGCCATCTCTTACCAGGCA 956
|
|
|
Db 1293 GAACAAAGTGTGAGTGAACAGCACCCGAGTACTGGGACACTGCCATCTCTTACCAGGCA 1352
|
|
|
QY 957 GGACTGTGTGGAGCCACCACTGTGACACCTTGGGATGCTGGGATGCTGGTACCATGTG 1016
|
|
|
Db 1353 GGACTGTGTGGAGCCACCACTGTGACACCTTGGGATGCTGGGATGCTGGTACCATGTG 1412
|
|
|
QY 1017 TGACCCCAAGAGAGTGTCTGTCTATTGAGGAGTGGCTTCCATTCAGCTTACCAC 1076
|
|
|
Db 1413 TGACCCCAAGAGAGTGTCTGTCTATTGAGGAGTGGCTTCCATTCAGCTTACCAC 1472
|
|
|
QY 1077 TGCCACAGCTGGGCGCACTGTTCACATGCCCCCATGACAATGTGAAAGTCTGTGAGGA 1136
|
|
|
Db 1473 TGCCACAGCTGGGCGCACTGTTCACATGCCCCCATGACAATGTGAAAGTCTGTGAGGA 1532
|
|
|
QY 1137 GGTGTTTGGGAAGTCCGAGCCCAACACATGATGTCCCCGACCTCATCCAGATCGACCG 1196
|
|
|
Db 1533 GGTGTTTGGGAAGTCCGAGCCCAACACATGATGTCCCCGACCTCATCCAGATCGACCG 1592
|
|
|
QY 1197 TGCCAAACCCCTGTGTCAGCTGCAAGTGGTCCCATCATCAGGACTTCTTGACAGCGGCA 1256
|
|
|
Db 1593 TGCCAAACCCCTGTGTCAGCTGCAAGTGGTCCCATCATCAGGACTTCTTGACAGCGGCA 1652
|
|
|
QY 1257 CGGTGACTGCTCTTGGACCAACCCAGCAAGCCATCTCCCTGCCGAGGATCTGCCGG 1316
|
|
|
Db 1653 CGGTGACTGCTCTTGGACCAACCCAGCAAGCCATCTCCCTGCCGAGGATCTGCCGG 1712
|
|
|
QY 1317 CGCCAGCTACACCTGAGCCAGCAGTGGCTTTTGGCGTGGGTTCACAGGCTCTG 1376
|
|
|
Db 1713 CGCCAGCTACACCTGAGCCAGCAGTGGCTTTTGGCGTGGGTTCACAGGCTCTG 1772
|
|
|
QY 1377 TCCTTACATGCACTACTGACACCAAGTGTGGTGCACCGGGAAGGCAAGGACAGATGT 1436
|
|
|
Db 1773 TCCTTACATGCACTACTGACACCAAGTGTGGTGCACCGGGAAGGCAAGGACAGATGT 1832
|
|
|
QY 1437 GTGCCAGACCGGCTTCCCTGGCGGCGATGCGACCACTGTGGGAGGCAAGCTCTG 1496
|
|
|
Db 1833 GTGCCAGACCGGCTTCCCTGGCGGCGATGCGACCACTGTGGGAGGCAAGCTCTG 1892
|
|
|

```

```
QY 1497 CCTCAAGGGGCTGGTGGAGAGACACACCTCAACAGCACAGGGTGGATGCTCTG 1556
Db 1893 CCTCAAGGGGCTGGTGGAGAGACACACCTCAACAGCACAGGGTGGATGCTCTG 1952
QY 1557 GGCCTAATGGATCCCTATATGGCCCTGCTCGGACACATGTGGGGGGCTGCAGCTGC 1616
Db 1953 GGCCTAATGGATCCCTATATGGCCCTGCTCGGACACATGTGGGGGGCTGCAGCTGC 2012
QY 1617 CAGGAGCAGTGCACCAACCCACCCCTGCACACGGGGGCAAGTACTGGAGGGAGTGAG 1676
Db 2013 CAGGAGCAGTGCACCAACCCACCCCTGCACACGGGGGCAAGTACTGGAGGGAGTGAG 2072
QY 1677 GGTGAATACCGATCTGCAACTGTGAGCCCTGAGCCCTGCCCCAGCTCAGCCCTCCGGAAGAGCTT 1736
Db 2073 GGTGAATACCGATCTGCAACTGTGAGCCCTGCCCCAGCTCAGCCCTCCGGAAGAGCTT 2132
QY 1737 CCGGGAGGAGCAGTGTGAGGCTTTCAACGCTTCAACACAGCACAGCAACCCGCTCACTCT 1796
Db 2133 CCGGGAGGAGCAGTGTGAGGCTTTCAACGCTTCAACACAGCACAGCAACCCGCTCACTCT 2192
QY 1797 CGCGTGGCATGGTGGCCCAAGTACTCCGCGGTGCTCCCGGGGACCAAGTCAAGCTCAT 1856
Db 2193 CGCGTGGCATGGTGGCCCAAGTACTCCGCGGTGCTCCCGGGGACCAAGTCAAGCTCAT 2252
QY 1857 CTGCCGAGCAATGGCACTGGCTACTTCTATGTGCTGGCACCAAGGTGGTGGACGGCAC 1916
Db 2253 CTGCCGAGCAATGGCACTGGCTACTTCTATGTGCTGGCACCAAGGTGGTGGACGGCAC 2312
QY 1917 GCTGTGCTCTCTGACTCCACTCCGCTGCTGTGTCCAAAGGAAGTGCATCAAGCTGGCTG 1976
Db 2313 GCTGTGCTCTCTGACTCCACTCCGCTGCTGTGTCCAAAGGAAGTGCATCAAGCTGGCTG 2372
QY 1977 TGATGGAACTCGGCTCCAAAGAGATTCGACAAAGTGTGGGTGTGGGGGAGACAA 2036
Db 2373 TGATGGAACTCGGCTCCAAAGAGATTCGACAAAGTGTGGGTGTGGGGGAGACAA 2432
QY 2037 TAAGAGTCAAGAAGTGTGACTGGCTCTTCAACAGCCCATGCTGGCTACAAATTTGCT 2096
Db 2433 TAAGAGTCAAGAAGTGTGACTGGCTCTTCAACAGCCCATGCTGGCTACAAATTTGCT 2492
QY 2097 GGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGCGGCTTACAAAGGCT 2156
Db 2493 GGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGCGGCTTACAAAGGCT 2552
QY 2157 GATCGGGATGACAACCTACCTGCTCTGAAGAAGCAGCAAGCAAGTACCTGCTCAACGG 2216
Db 2553 GATCGGGATGACAACCTACCTGCTCTGAAGAAGCAGCAAGCAAGTACCTGCTCAACGG 2612
QY 2217 GCATTTCTGCTGCTGGCGGTGGAGCGGACCTGGTGGTGAAGGCAAGTCTGTCGGTA 2276
Db 2613 GCATTTCTGCTGCTGGCGGTGGAGCGGACCTGGTGGTGAAGGCAAGTCTGTCGGTA 2672
QY 2277 CAGCGGACGGGACACGCTGGAGAGCCCTGCAGGCTTCCCGGCCATCTCGGAGCGCT 2336
Db 2673 CAGCGGACGGGACACGCTGGAGAGCCCTGCAGGCTTCCCGGCCATCTCGGAGCGCT 2732
QY 2337 GACCGTGGAGTCTCTCCGCTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTA 2396
Db 2733 GACCGTGGAGTCTCTCCGCTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTA 2792
QY 2397 TCTGCCAAGAGGCTCGGAGGAGCAAGTCTCTCATCCCAAGGACCCCGGGGACCCCTC 2456
Db 2793 TCTGCCAAGAGGCTCGGAGGAGCAAGTCTCTCATCCCAAGGACCCCGGGGACCCCTC 2852
QY 2457 TGTCTTGACAAACAGCTCTCAGCCCTCTCAACAGGTGGAGAGCCGACGACAGGCG 2516
Db 2853 TGTCTTGACAAACAGCTCTCAGCCCTCTCAACAGGTGGAGAGCCGACGACAGGCG 2912
QY 2517 CCCTGCACGCTGGGTGGCTGGCAGCTGGGGCGCTGCTCCGCGAGCTGGCGGAGTGCCT 2576
Db 2913 CCCTGCACGCTGGGTGGCTGGCAGCTGGGGCGCTGCTCCGCGAGCTGGCGGAGTGCCT 2972
```

```
QY 2577 GCAGAAGCGGCGGTGGAGTCCCGGGGCTCCGCGGCGACGGCAGCGTCCCTGCCTGTGA 2636
Db 2973 GCAGAAGCGGCGGTGGAGTCCCGGGGCTCCGCGGCGACGGCAGCGTCCCTGCCTGTGA 3032
QY 2637 TGCAGCCCATCGGCGCTGGAGACACAAGCTCGGGGAGCCCTGCCCACTCGGAGCT 2696
Db 3033 TGCAGCCCATCGGCGCTGGAGACACAAGCTCGGGGAGCCCTGCCCACTCGGAGCT 3092
QY 2697 CAGGCGCTGGTCAACCTGCTCAAGAGTGGCGCGGGGATTTTCAGAGGGGCTCACTCAA 2756
Db 3093 CAGGCGCTGGTCAACCTGCTCAAGAGTGGCGCGGGGATTTTCAGAGGGGCTCACTCAA 3152
QY 2757 GTGTGTGGGCGCACAGGAGCGGCTGCTGGCGCGGGACAGTGCAACTTGCACCGCAAGCC 2816
Db 3153 GTGTGTGGGCGCACAGGAGCGGCTGCTGGCGCGGGACAGTGCAACTTGCACCGCAAGCC 3212
QY 2817 CCAGGAGCTGAGCTTCTGCGTCCCTGAGCGCGTCTGA 2853
Db 3213 CCAGGAGCTGAGCTTCTGCGTCCCTGAGCGCGTCTGA 3249

RESULT 3
US-10-163-316-3
; Sequence 3, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MP101-025P1RNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2469)
US-10-163-316-3
```

Query Match 50.3%; Score 1435; DB 9; Length 2469;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1485; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 57 TCAGCCAGAGCGGAGGTAGTCTCCATCGACTGGACCGGGACATTAACGGCGCGG 116
Db 57 TCAGCCAGAGCGGAGGTAGTCTCCATCGACTGGACCGGGACATTAACGGCGCGG 116
QY 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTLAGATCAGC 176
Db 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTLAGATCAGC 176
QY 177 ATTTACAGGAGACTTTTACCTACCTGACCGCGGATCTCAGTTCTTTGGCTCCCGCTT 236
Db 177 ATTTACAGGAGACTTTTACCTACCTGACCGCGGATCTCAGTTCTTTGGCTCCCGCTT 236
QY 237 CTCCACTGAGCATCTGGGCGTCCCGTCCAGGGGCTCACCAGGGGCTCTTCAGACTGCG 296
Db 237 CTCCACTGAGCATCTGGGCGTCCCGTCCAGGGGCTCACCAGGGGCTCTTCAGACTGCG 296
QY 297 AGCTGCTTCTATTTCTGGGGAGTGAACCCGAGCGGAGTGTTCGCTGCTGAGCCT 356
Db 297 AGCTGCTTCTATTTCTGGGGAGTGAACCCGAGCGGAGTGTTCGCTGCTGAGCCT 356
QY 357 GTGCGGGGGCTCCGCGAGGCTTTTGGCTACCGAGCGCGGAGTGTTCATTTAGCCCGCT 416
Db 357 GTGCGGGGGCTCCGCGAGGCTTTTGGCTACCGAGCGCGGAGTGTTCATTTAGCCCGCT 416
```

QY 417 GCCCAATGCTAGCGCGCGCGCGAGCGCAAGAGCCAGGCGCACACCTCTCCAGCG 476
Db 417 GCCCAATGCTAGCGCGCGCGCGAGCGCAAGAGCCAGGCGCACACCTCTCCAGCG 476
QY 477 CCGGGGTGTTCCGGGCGGGGCTTCCGGAGACCCCACTCTCGCTGCGGGGTGGCCTCGG 536
Db 477 CCGGGGTGTTCCGGGCGGGGCTTCCGGAGACCCCACTCTCGCTGCGGGGTGGCCTCGG 536
QY 537 CTGGAACCCCGCCATCTACGGGCGCTTGGACCTTACAGCCGCGGCGCGGCTTCGG 596
Db 537 CTGGAACCCCGCCATCTACGGGCGCTTGGACCTTACAGCCGCGGCGCGGCTTCGG 596
QY 597 GGAGAGTCGTAGCGGCGCGAGTCTGGGCGCGCAAGCGCTTCTGCTATCCCGGGTA 656
Db 597 GGAGAGTCGTAGCGGCGCGAGTCTGGGCGCGCAAGCGCTTCTGCTATCCCGGGTA 656
QY 657 CGTGAGACGCTGCTGCTGCGGCGCGAGTCAAGTTCAGGTCACGCGCGGACCTGGA 716
Db 657 CGTGAGACGCTGCTGCTGCGGCGCGAGTCAAGTTCAGGTCACGCGCGGACCTGGA 716
QY 717 ACATTATCTGCTGACGCTGCTGCAACGCGGCGGCGACTACCGGCATCCCGAGCATCT 776
Db 717 ACATTATCTGCTGACGCTGCTGCAACGCGGCGGCGACTACCGGCATCCCGAGCATCT 776
QY 777 CAACCCCATCAACATGCTGCTGAAGTGTCTTCTTAGAGTCGTGACTCCGGGCC 836
Db 777 CAACCCCATCAACATGCTGCTGAAGTGTCTTCTTAGAGTCGTGACTCCGGGCC 836
QY 837 CAAGGTCACCGGCAATGCGGCGCTGACGCTGGCGCACTTCTGCTGGCGAAGAGCT 896
Db 837 CAAGGTCACCGGCAATGCGGCGCTGACGCTGGCGCACTTCTGCTGGCGAAGAGCT 896
QY 897 GACAAAGTAGTGACAAAGACCCGAGTACTGGGACACTGGCATCTCTTCCACAGGCA 956
Db 897 GACAAAGTAGTGACAAAGACCCGAGTACTGGGACACTGGCATCTCTTCCACAGGCA 956
QY 957 GGACCTGTGTGGAGCACCACCTGTGACACCTTGGCGATGGCTGATGGGTACCATGTG 1016
Db 957 GGACCTGTGTGGAGCACCACCTGTGACACCTTGGCGATGGGTGATGGGTACCATGTG 1016
QY 1017 TGACCCCAAGAGAGAGTGTCTGTCAATTGAGGACATGGGCTTCCATCAGCCTTCCACAC 1076
Db 1017 TGACCCCAAGAGAGTGTCTGTCAATTGAGGACATGGGCTTCCATCAGCCTTCCACAC 1076
QY 1077 TGCCACGAGCTGGGCGAGCTGTTCAATGCGCCCATGACAAATGTGAAGTCTGTGAGGA 1136
Db 1077 TGCCACGAGCTGGGCGAGCTGTTCAATGCGCCCATGACAAATGTGAAGTCTGTGAGGA 1136
QY 1137 GGTGTTTGGGAAGTCCGAGCAACCATGATGTCCCGACCTCATCCAGATCGACCG 1196
Db 1137 GGTGTTTGGGAAGTCCGAGCAACCATGATGTCCCGACCTCATCCAGATCGACCG 1196
QY 1197 TGCCAAACCCCTGGTACGCTGCAAGTGTGCTGCAATCATCAGGACTTCTGGACAGCGGCA 1256
Db 1197 TGCCAAACCCCTGGTACGCTGCAAGTGTGCTGCAATCATCAGGACTTCTGGACAGCGGCA 1256
QY 1257 CGGTGACTGCTTGGACCAACCCAGCCCATCTCCCTGTCGCCGAGGATCTGCCGGG 1316
Db 1257 CGGTGACTGCTTGGACCAACCCAGCCCATCTCCCTGTCGCCGAGGATCTGCCGGG 1316
QY 1317 CGCCAGCTACACCTTGAGCAGCAGTGGGAGTGGCTTGGCGTGGCTTCCAGCCCTG 1376
Db 1317 CGCCAGCTACACCTTGAGCAGCAGTGGGAGTGGCTTGGCGTGGCTTCCAGCCCTG 1376
QY 1377 TCCTTACATGCACTGCAACAGCTGTGTGTCACCGGAGGCGCAAGGACAGATGCT 1436
Db 1377 TCCTTACATGCACTGCAACAGCTGTGTGTCACCGGAGGCGCAAGGACAGATGCT 1436
QY 1437 GTGCCAGACCCGCACTTCCCTGGGCGATGGCAGCAGCTGTGGCGAGGCGAGCTCTG 1496
Db 1437 GTGCCAGACCCGCACTTCCCTGGGCGATGGCAGCAGCTGTGGCGAGGCGAGCTCTG 1496
QY 1497 CCTAAAGGGGCGCTGCGTGGAGAGACACAACCTCAACAGCACAGG 1542

Db 1497 CCTAAAGGGGCGCTGCGTGGAGAGACACAACCTCAACAGCACAGG 1542

RESULT 4

US-10-163-316-1
; Sequence 1, Application US/10163316
; Publication No. US2002019703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: MP101-025P1RNM
; CURRENT APPLICATION NUMBER: US/10163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2940
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (472)...(2941)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2940)
; OTHER INFORMATION: n = A,T,C or G
US-10-163-316-1

Query Match 50.3%; Score 1435; DB 9; Length 2940;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1485; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 57 TGAGCCAGAGCGGGAGGTAGTCTCCATCCGACTCCGAGCGGACATTAACGCGCGCG 116
Db 528 TGAGCCAGAGCGGGAGGTAGTCTCCATCCGACTCCGAGCGGACATTAACGCGCGCG 587
QY 117 CTACTACTGCGGGGTCCCGAGGACTCGGGGATCAGGGACTCAATTTTCAATCACAGC 176
Db 588 CTACTACTGCGGGGTCCCGAGGACTCGGGGATCAGGGACTCAATTTTCAATCACAGC 647
QY 177 ATTTACAGGAGACTTTTACCTACACCTGACCGCGGATGCTAGTCTTGGCTCCCGCTT 236
Db 648 ATTTACAGGAGACTTTTACCTACACCTGACCGCGGATGCTAGTCTTGGCTCCCGCTT 707
QY 237 CTCACATGAGCATCTGGGCGCTCCCGCTCCAGGGGCTCCAGGGGCTCTTTCAGACTGCG 296
Db 708 CTCACATGAGCATCTGGGCGCTCCCGCTCCAGGGGCTCCAGGGGCTCTTTCAGACTGCG 767
QY 297 ACGTGTCTTCTATTTCTGGGACGTGAACGCGGAGCGGACTGCTTCCGCTGTGAGCCT 356
Db 768 ACGTGTCTTCTATTTCTGGGCGGTGAACGCGGAGCGGACTGCTTCCGCTGTGAGCCT 827
QY 357 GTGCGGGGGCTCCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGCTATTAGCCGCT 416
Db 828 GTGCGGGGGCTCCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGCTATTAGCCGCT 887
QY 417 GCCCAATGCTAGCGCGCGCGGCGAGCGCAACAGCAGGCGGCGACACCTTCTCCAGCG 476
Db 888 GCCCAATGCTAGCGCGCGCGGCGAGCGCAACAGCAGGCGGCGACACCTTCTCCAGCG 947
QY 477 CCGGGGTGTTCCGGGCGGGCTTCCGAGAGCCCTTCCGCTGCGGGGTGGCTTCGGG 536
Db 948 CCGGGGTGTTCCGGGCGGGCTTCCGAGAGCCCTTCCGCTGCGGGGTGGCTTCGGG 1007
QY 537 CTGGAACCCCGCCATCTTACGGGCGCTTGGACCTTACAAAGCCGCGCGGCTTCGG 596
Db 1008 CTGGAACCCCGCCATCTTACGGGCGCTTGGACCTTACAAAGCCGCGGCGGCTTCGG 1067
QY 597 GGAGAGTCGTAGCGGCGCGAGTCTGGGCGCGCAAGCGTTCGTGCTATCCCGGGTA 656

```
|||||
Db 1068 GGAGAGTCGTAGCCGGCGCAGGCTCTGGCGCGCAAGCGTTGCTCTATCCCGGGTA 1127
QY 657 CGTGGAGACGCTGGTGGTCCGCGACGAGTCAATGGTCAAGTTCACCGCGCGGACCTGGA 716
Db 1128 CGTGGAGACGCTGGTGGTCCGCGACGAGTCAATGGTCAAGTTCACCGCGCGGACCTGGA 1187
QY 717 ACATTATCTGCTGACGCTGCTGCAACGCGCGCGGCGGACCTCTACCGCATCCAGCATCCT 776
Db 1188 ACATTATCTGCTGACGCTGCTGCAACGCGCGCGGCGGACCTCTACCGCATCCAGCATCCT 1247
QY 777 CAACCCCATCAACATCTGTTGGTCAAGTGCCTCTTTAGAGATCGTGACTCCGGGCC 836
Db 1248 CAACCCCATCAACATCTGTTGGTCAAGTGCCTCTTTAGAGATCGTGACTCCGGGCC 1307
QY 837 CAAGGTCAACGGGCAATCGGCCCTGACGCTGCGCAACTCTGTGCTTGGCGAGAGACT 896
Db 1308 CAAGGTCAACGGGCAATCGGCCCTGACGCTGCGCAACTCTGTGCTTGGCGAGAGACT 1367
QY 897 GAACAAAGTGAGTGACAAAGCACCCGAGTACTGGGACACTGCCATCCTCTTCAACGAGCA 956
Db 1368 GAACAAAGTGAGTGACAAAGCACCCGAGTACTGGGACACTGCCATCCTCTTCAACGAGCA 1427
QY 957 GGACCTGTGTGGAGCCACCACTGTGACACCTTGGGATGGCTGTGATGGGTACCATGTG 1016
Db 1428 GGACCTGTGTGGAGCCACCACTGTGACACCTTGGGATGGCTGTGATGGGTACCATGTG 1487
QY 1017 TGACCCCAAGAGAGCTGCTGTGCTATGAGGAGGATGGCTTCCATCAGCCTTCAACCAC 1076
Db 1488 TGACCCCAAGAGAGCTGCTGTGCTATGAGGAGGATGGCTTCCATCAGCCTTCAACCAC 1547
QY 1077 TGCCACAGAGCTGGGCGCACGCTGTTCAACATGCCCATGACAAATGTGAAAGTCTGTGAGGA 1136
Db 1548 TGCCACAGAGCTGGGCGCACGCTGTTCAACATGCCCATGACAAATGTGAAAGTCTGTGAGGA 1607
QY 1137 GGTGTTGGGAAGCTCCGAGGCAACACATGATGTCCCGACCTCATCAGATCGACCG 1196
Db 1608 GGTGTTGGGAAGCTCCGAGGCAACACATGATGTCCCGACCTCATCAGATCGACCG 1667
QY 1197 TGCCAAACCCCTGTGACGCTGCACTGTGTCATCATCAGGACTTCCGTGACAGCGGGCA 1256
Db 1668 TGCCAAACCCCTGTGACGCTGCACTGTGTCATCATCAGGACTTCCGTGACAGCGGGCA 1727
QY 1257 CGGTGACTGCTCTGTGACCAACCCAGACCCATCCTGCTGCGCCAGGATCTGCCGG 1316
Db 1728 CGGTGACTGCTCTGTGACCAACCCAGACCCATCCTGCTGCGCCAGGATCTGCCGG 1787
QY 1317 CGCAGCTACACCTGAGCCAGCAGTGCAGCTGCTGCTGCTGGGCTCCAGGCTG 1376
Db 1788 CGCAGCTACACCTGAGCCAGCAGTGCAGCTGCTGCTGCTGGGCTCCAGGCTG 1847
QY 1377 TCCTTACATGAGTACTGCACCAAGCTGTGGTGCACCGGAGGCCAAGGACAGATGGT 1436
Db 1848 TCCTTACATGAGTACTGCACCAAGCTGTGGTGCACCGGAGGCCAAGGACAGATGGT 1907
QY 1437 GTGCCAGACCCGCACTTCCCTCGGCGGATGSCACAGCTGTGCGGAGGSCAGCTCTG 1496
Db 1908 GTGCCAGACCCGCACTTCCCTCGGCGGATGSCACAGCTGTGCGGAGGSCAGCTCTG 1967
QY 1497 CCTCAAGGGGCTGCTGGGAGACACAACTCTCAACAGCACAGG 1542
Db 1968 CCTCAAGGGGCTGCTGGGAGACACAACTCTCAACAGCACAGG 2013
```

RESULT 5
US-09-965-631-5
; Sequence 5, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION: Carl Johan
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Encodin
; FILE REFERENCE: LEX-0241-USA

```
; CURRENT APPLICATION NUMBER: US/09/965,631  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: US 60/236,689  
; PRIOR FILING DATE: 2000-09-29  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1104  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-09-965-631-5
```

Query Match 36.3%; Score 1035; DB 10; Length 1104;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1035; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 57 TCAGCCACAGCGGAGGTAGTCTCCATCGACTGACCGGACATTAACGGCGCGCG 116
Db 57 TCAGCCACAGCGGAGGTAGTCTCCATCGACTGACCGGACATTAACGGCGCGCG 116
QY 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTTCAGATCACAGC 176
Db 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTTCAGATCACAGC 176
QY 177 ATTTTCAGAGGAGCTTTTACCTACACTGACCCGGATGCTCAGTTCTTGGCTCC-GCCCTT 236
Db 177 ATTTTCAGAGGAGCTTTTACCTACACTGACCCGGATGCTCAGTTCTTGGCTCC-GCCCTT 236
QY 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCCACGGGGGCTCTTCAGACCTCGG 296
Db 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCCACGGGGGCTCTTCAGACCTCGG 296
QY 297 ACGTGTCTTATCTTGGGGACGTGAACCGCCGAGCGGACTGTTGCTGCTGCTGAGGCT 356
Db 297 ACGTGTCTTATCTTGGGGACGTGAACCGCCGAGCGGACTGTTGCTGCTGCTGAGGCT 356
QY 357 GTGCGGGGGGCTCCGCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCCGCT 416
Db 357 GTGCGGGGGGCTCCGCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCCGCT 416
QY 417 GCCCAATGCTAGCGCGCGCGCGAGCGCAACAGCAGGCGGACACCTTCTCCAGGG 476
Db 417 GCCCAATGCTAGCGCGCGCGCGAGCGCAACAGCAGGCGGACACCTTCTCCAGGG 476
QY 477 CGGGGTGTTCCGGGCGGGGCTTCCGGAGACCCACCTCTCCGCTCGGGGTGCGCTCGGG 536
Db 477 CGGGGTGTTCCGGGCGGGGCTTCCGGAGACCCACCTCTCCGCTCGGGGTGCGCTCGGG 536
QY 537 CTGGAACCCCGCATCTTACGGGCGCTTGACACCTTTACAAGCCGCGGCGGGGCTTCGG 596
Db 537 CTGGAACCCCGCATCTTACGGGCGCTTGACACCTTTACAAGCCGCGGCGGGGCTTCGG 596
QY 597 GGAGAGTCGTAGCGCGCGAGGTGCGGGCGCGCAAGCGTTTCTGCTATATCCCGGGTA 656
Db 597 GGAGAGTCGTAGCGCGCGAGGTGCGGGCGCGCAAGCGTTTCTGCTATATCCCGGGTA 656
QY 657 CGTGGAGACGCTGCTGCTCGGACAGTCAATGGTCAAGTTCACAGCGCGGAGCTGGA 716
Db 657 CGTGGAGACGCTGCTGCTCGGACAGTCAATGGTCAAGTTCACAGCGCGGAGCTGGA 716
QY 717 ACATTATCTGCTGACGCTGCTGGCAACGCGGCGGAGCTCTACCGCATCCAGCATCCT 776
Db 717 ACATTATCTGCTGACGCTGCTGGCAACGCGGCGGAGCTCTACCGCATCCAGCATCCT 776
QY 777 CAACCCCATCAACATCTGTTGGTCAAGTGTGCTTCTTAGAGATCGTGACTCCGGGCC 836
Db 777 CAACCCCATCAACATCTGTTGGTCAAGTGTGCTTCTTAGAGATCGTGACTCCGGGCC 836
QY 837 CAAGGTCAACGGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGCGAGAGACT 896
Db 837 CAAGGTCAACGGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGCGAGAGACT 896
QY 897 GAACAAAGTGAGTGACAAAGCACCCGAGTACTGGGACACTGCCATCCTCTTCAACGAGCA 956
```

Db 897 GAACAAAGTGTAGTGACAAAGCACCACCGAGTACTGGGACACTGCGCATCTCTTACACAGGCA 956
 QY 957 GGACCTGTGTGGAGCCACCACTGTGTACACCCCTGGGATGGCTGTGGTACCATGTG 1016
 Db 957 GGACCTGTGTGGAGCCACCACTGTGTACACCCCTGGGATGGCTGTGGTACCATGTG 1016
 QY 1017 TGACCCCAAGAGAAGCTGCTGTGTATTTAGAGACGATGGGCTTCCATCATCAGCCCTTACAC 1076
 Db 1017 TGACCCCAAGAGAAGCTGCTGTGTATTTAGAGACGATGGGCTTCCATCATCAGCCCTTACAC 1076
 QY 1077 TGCCACAGCTGGG 1091
 Db 1077 TGCCACAGCTGGG 1091

RESULT 6
 US-09-965-631-1
 ; Sequence 1, Application US/09965631
 ; Patent No. US20020115842A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fridgele, Carl Johan
 ; APPLICANT: Hilbun, Erin
 ; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Encodin
 ; FILE REFERENCE: LEX-0241-USA
 ; CURRENT APPLICATION NUMBER: US/09/965,631
 ; CURRENT FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: US 60/236,689
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 966
 ; TYPE: DNA
 ; ORGANISM: homo sapiens
 ; US-09-965-631-1

Query Match 31.7% Score 903; DB 10; Length 966;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 903; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 57 TGAGCCAGAGCGGAGGTAGTTCCTCCATCCGACTGACCGCGGACATTAACGCGCGCG 116
 Db 57 TGAGCCAGAGCGGAGGTAGTTCCTCCATCCGACTGACCGCGGACATTAACGCGCGCG 116
 QY 117 CTACTACTGGGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAGC 176
 Db 117 CTACTACTGGGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAGC 176
 QY 177 ATTTTCAGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGCTCCGCGCTT 236
 Db 177 ATTTTCAGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGCTCCGCGCTT 236
 QY 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGCG 296
 Db 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGCG 296
 QY 297 AGCTGCTTCTATTCTGGGGAGGTGAACCGCGAGCGGACTGCTGCTGCTGTGAGCCT 356
 Db 297 AGCTGCTTCTATTCTGGGGAGGTGAACCGCGAGCGGACTGCTGCTGCTGTGAGCCT 356
 QY 357 GTGCGGGGCTCCCGGAGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCCGCT 416
 Db 357 GTGCGGGGCTCCCGGAGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCCGCT 416
 QY 417 GCCCAATGTAGCGCGCGGCGGCGAGCGCAACAGCAGGCGCACACCTTCTCCAGCG 476
 Db 417 GCCCAATGTAGCGCGCGGCGGCGAGCGCAACAGCAGGCGCACACCTTCTCCAGCG 476
 QY 477 CCGGGGTGTCCGGGGGCTTCCGGAGACCCCGACCTCTCGCTCGGGGTGGCGCTCGG 536
 Db 477 CCGGGGTGTTCGGGGGCGGCTTCCGGAGACCCCGACCTCTCGCTCGGGGTGGCGCTCGG 536

QY 537 CTGGAACCCCGCCCATCTACGGGCGCTTGAACCTTACAAGCCCGCGCGGCGGCTTCGG 596
 Db 537 CTGGAACCCCGCCCATCTACGGGCGCTTGAACCTTACAAGCCCGCGCGGCGGCTTCGG 596
 QY 597 GGAGAGTCTAGCCGCGCAGGTCTGGGCGCGCAAGCGTTTCTGTCTATCCCGCGGTA 656
 Db 597 GGAGAGTCTAGCCGCGCAGGTCTGGGCGCGCAAGCGTTTCTGTCTATCCCGCGGTA 656
 QY 657 CGTGAGAGCGCTGGTGTCTCGGACGAGTCAATGGTCAAGTTCACGCGCGGAGCTGGA 716
 Db 657 CGTGAGAGCGCTGGTGTCTCGGACGAGTCAATGGTCAAGTTCACGCGCGGAGCTGGA 716
 QY 717 ACATTATCTGTGACGCTGTGTCGCAACGCGCGGCGGACTTACCGCATCCAGCATCT 776
 Db 717 ACATTATCTGTGACGCTGTGTCGCAACGCGCGGCGGACTTACCGCATCCAGCATCT 776
 QY 777 CAACCCCATCAACATCTGTGTGTCGTCGTCCTTTAGAGATCGTGACTTCCGGGCG 836
 Db 777 CAACCCCATCAACATCTGTGTGTCGTCGTCCTTTAGAGATCGTGACTTCCGGGCG 836
 QY 837 CAAGGTACCGGCAATGCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCNGAGAAGCT 896
 Db 837 CAAGGTACCGGCAATGCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCNGAGAAGCT 896
 QY 897 GAACAAAGTGTAGTGACAAAGCACCACCGGAGTACTGGGACACTGCCATCTCTTCAACAGGCA 956
 Db 897 GAACAAAGTGTAGTGACAAAGCACCACCGGAGTACTGGGACACTGCCATCTCTTCAACAGGCA 956
 QY 957 GGA 959
 Db 957 GGA 959

RESULT 7
 US-10-105-929-7
 ; Sequence 7, Application US/10105929
 ; Patent No. US20020137142A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Holtzman, Douglas A.
 ; APPLICANT: Goodearl, Andrew D.J.
 ; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
 ; FILE REFERENCE: 09404/041001
 ; CURRENT APPLICATION NUMBER: US/10/105,929
 ; CURRENT FILING DATE: 2002-03-25
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 7
 ; LENGTH: 2114
 ; TYPE: DNA
 ; ORGANISM: Rattus rattus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (3)...(1445)
 ; US-10-105-929-7

Query Match 0.8% Score 23; DB 12; Length 2114;
 Best Local Similarity 100.0%; Pred. No. 0.078;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2008 GACAAGTGTGGGTGTGGGGG 2030
 Db 807 GACAAGTGTGGGTGTGGGGG 829

RESULT 8
 US-09-931-457A-10
 ; Sequence 10, Application US/09931457A

```

; Patent No. US20020157132A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes
; FILE REFERENCE: BB1116 US CIP
; CURRENT APPLICATION NUMBER: US/09/931,457A
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 09/424,976
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/065,385
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/049,406
; PRIOR FILING DATE: 1997-06-12
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 1813
; TYPE: DNA
; ORGANISM: Zea mays
; US-09-931-457A-10

Query Match      0.8%; Score 22; DB 9; Length 1813;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1659 GTACTGCGAGGAGTGAGGCTG 1680
Db 269 GTACTGCGAGGAGTGAGGCTG 290

RESULT 9
US-09-918-995-22019/c
; Sequence 22019, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 05/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22019
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(506)
; OTHER INFORMATION: n = A,T,C or G
; US-09-918-995-22019

Query Match      0.7%; Score 21; DB 9; Length 506;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1608 GCAGCTGGCCAGGAGGAGTG 1628
Db 346 GCAGCTGGCCAGGAGGAGTG 326

RESULT 10
US-09-925-297-180/c
; Sequence 180, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105

; Patent No. US20020157132A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes
; FILE REFERENCE: BB1116 US CIP
; CURRENT APPLICATION NUMBER: US/09/931,457A
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 09/424,976
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/065,385
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/049,406
; PRIOR FILING DATE: 1997-06-12
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 1827
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1524)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-297-180

Query Match      0.7%; Score 21; DB 10; Length 1827;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1608 GCAGCTGGCCAGGAGGAGTG 1628
Db 1606 GCAGCTGGCCAGGAGGAGTG 1586

RESULT 11
US-09-938-842A-60
; Sequence 60, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI3300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 60
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-60

Query Match      0.7%; Score 20; DB 9; Length 1017;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1996 AAGAAGAGATTGCAAGTG 2015
Db 468 AAGAAGAGATTGCAAGTG 487

RESULT 12
US-09-535-459-1835
; Sequence 1835, Application US/09535459
; Publication No. US20030040615A1
; GENERAL INFORMATION:
; APPLICANT: Sellmeier, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
```

```
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING ELECTRON TRANSFER MOLECULE
; FILE REFERENCE: PD-1014 CIP
; CURRENT APPLICATION NUMBER: US/09/535,459
; CURRENT FILING DATE: 2000-03-24
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2170
; SOFTWARE: PERL Program
; SEQ ID NO 1835
; LENGTH: 256
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030040615A1 rat00012080
; NAME/KEY: unsure
; LOCATION: 205-206, 221
; OTHER INFORMATION: a, t, c, g, or other
US-09-535-459-1835

Query Match          0.7%; Score 19; DB 9; Length 256;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 889 AAGAGCTCAACAAAGTGA 907
Db 238 AAGAGCTCAACAAAGTGA 256
|||||

RESULT 13
US-09-918-995-15962/c
; Sequence 15962, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15962
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-15962

Query Match          0.7%; Score 19; DB 9; Length 408;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2678 CCTGCCCCACCTGGGAGCT 2696
Db 287 CCTGCCCCACCTGGGAGCT 269
|||||

RESULT 14
US-09-925-297-85
; Sequence 85, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
```

```
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 431
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (325)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (334)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (347)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (381)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-85

Query Match          0.7%; Score 19; DB 10; Length 431;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 TCCGGGCCCAAGGTACCG 847
Db 44 TCCGGGCCCAAGGTACCG 62
|||||

RESULT 15
US-09-918-995-32286
; Sequence 32286, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32286
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-32286

Query Match          0.7%; Score 19; DB 9; Length 452;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1709 GCCCCAGCTCAGCTCCGG 1727
Db 384 GCCCCAGCTCAGCTCCGG 402
|||||

Search completed: May 2, 2003, 01:19:03
Job time : 382 secs
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 01:01:17 ; Search time 3889 seconds
(without alignments)
11881.132 Million cell updates/sec

Title: US-10-009-332-2

Perfect score: 2853
Sequence: 1 atgctttgtggcattct.....gcgtctgagccgtgctga 2853

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1:	em_estba:*
2:	em_esthum:*
3:	em_estlin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hic:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_hic:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pln:*
21:	em_gss_vit:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	53	1.9	544	12 BF078689	BF078689 229090 MA
c 2	39	1.4	195	10 AW935807	AW935807 QV3-DT001
c 3	38	1.3	528	17 A0809642	A0809642 HS.4758_A
c 4	36	1.3	961	10 BB612189	BB612189 BB612189
c 5	35	1.2	469	17 A2842795	A2842795 2M014119
c 6	35	1.2	522	10 BE553572	BE553572 ur47b04.y

7	1.1	305	12	BF320986	BF320986 uz57hl0.y
8	1.0	400	17	AQ440250	AQ440250 HS_5071.B
9	1.0	415	17	AW485848	AW485848 68315 MAR
10	27	169	17	B01986	B01986 CSRL-145A3-
11	26	627	10	BE281680	BE281680 601099657
12	25	247	10	BB604671	BB604671 BB604671
13	25	849	13	BI103177	BI103177 602889417
14	24	464	10	BE666088	BE666088 149119 MA
15	24	664	13	BI692986	BI692986 603344521
16	23	681	13	BG921734	BG921734 602825569
17	23	1011	17	CNS0471P	AL277414 Tetraodon
18	22	281	10	BB195208	BB195208 BB195208
19	22	314	10	BB154064	BB154064 BB154064
20	22	577	9	AI770556	AI770556 606054C06
21	22	614	14	B0037780	B0037780 946142B02
22	22	649	10	AW061788	AW061788 660012A07
23	22	2737	11	AY109369	AY109369 Zea mays
24	21	109	9	AI423690	AI423690 tf85f08.x
25	21	123	17	CNS010SU	AL169231 Tetraodon
26	21	207	10	AW105581	AW105581 xg33h10.x
27	21	215	9	AA303391	AA303391 EST15998
28	21	221	9	AA599823	AA599823 ag09b04.s
29	21	233	9	AI270237	AI270237 qu83c05.x
30	21	235	10	BB075309	BB075309 BB075309
31	21	238	10	AW046359	AW046359 UI-M-BH1-
32	21	241	9	AI233548	AI233548 EST230236
33	21	244	9	AI449663	AI449663 mr52e05.x
34	21	245	10	AW130644	AW130644 xg64a07.x
35	21	255	9	AI202856	AI202856 q15ah11.x
36	21	256	12	BF283275	BF283275 EST447866
37	21	256	13	BI345593	BI345593 397660 MA
38	21	260	9	AI698894	AI698894 wc95f04.x
39	21	261	10	AW430043	AW430043 68812 MAR
40	21	278	9	AA996648	AA996648 UI-R-CO-h
41	21	283	12	BF511038	BF511038 UI-H-B14-
42	21	301	9	AA478616	AA478616 zv19c12.s
43	21	302	9	AI499824	AI499824 tm92h10.x
44	21	305	9	AI580450	AI580450 to34a06.x
45	21	311	10	BE645522	BE645522 7e19i08.x

ALIGNMENTS

RESULT 1
BF078689/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BF078689 544 bp mRNA linear EST 18-OCT-2000
229090 MARC 2PTG Sus scrofa cDNA 5', mRNA sequence.

BF078689.1 GI:10872519

EST.

pig.

Sus scrofa

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 544)

Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.F.L., Casas, E.,

Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.H., Laegreid, W.W.

and Keeler, J.W.

Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 18

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BASE COUNT	37 a	60 c	52 q	46 t
tissue mRNA and cDNA amplification were performed under low stringency conditions."				

	Query Match	1.48;	Score 39;	DB 10;	Length 195;
Best Local Similarity	100.0%;	Pred. No. 2.1e-07;			
Matches 39;	Conservative 0;	Mismatches 0;	Indels		
QY	1542	GGTGGATGTTCTCTGGGCCAAATGGATCCCTATGGGCC	1580		
db	155	GGTGGATGTTCTCTGGGCCAAATGGATCCCTATGGGCC	117		

[illegible]

SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 528)

AUTHORS	Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and Hood, L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	PROC. NATL. ACAD. SCI. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 4758 row: G column: 21
Seq primer: T7
Class: BAC ends

```

FEATURES
  source
    High quality sequence stop: 528.
    Location/Qualifiers
      1..528
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="plate:4758 Col-21 Row-G"
        /clone_lib="CIT Approved Human Genomic Sperm Library D"
        /sex="male"
        /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones In
          E-Coli DH10B"
      BASE COUNT      138 a      145 c      100 g      142 t      3 others
      ORIGIN
        Query Match      1.3%;      Score 38;      DB 17;      Length 528;
        Best Local Similarity 100.0%;      Pred. No. 7.9e-07;
        Matches 38;      Conservative
        Indels 0;      Gaps 0;

```

BB612189	961 bp	mRNA	linear	
BB612189	RIKEN full-length enriched,	0 day neonate skin Mus		
BB612189	musculus	CDNA clone 4631401M01 5', mRNA sequence.		
DEFINITION				
LOCUS				
BB612189				EST 26-OCT-2000

ACCESSION	BB612189
VERSION	BB612189.1
KEYWORDS	GI:16453236
SOURCE	EST.
ORGANISM	house mouse. Mus musculus Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 961) Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda, M., Koya,Y., Matsuyama,T., Miyazaki,A., Nomura,K., Ono,M., Okada, Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki, D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y. RIKEN Mouse ESTs (Arakawa,T., et al. 2001) Unpublished (2001) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata.K., Itoh.M., Carninci.P., Sugahara Y. and Hayashizaki.Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo.S., Shinagawa.A., Saito,T., Kiyosawa.H., Yamanaka,I., Aizawa K., Fukuda,S., Hara.A., Itoh.M., Kawai.J., Shibata.K. and Hayashizaki,Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues.
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
BASE COUNT	158 a 302 c 301 g 200 t
ORIGIN	
Query Match	1.3%; Score 36; DB 10; Length 961;
Best Local Similarity	100.0%; Pred. No. 8.3e-06;
Matches	36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	449 ACAGCGAGGGGCGCACACTTCTCCAGCGCGGGGTG 484 Db 768 ACAGCGAGGGGCGCACACTTCTCCAGCGCGGGGTG 803
RESULT 5	
AZB42795/c	
LOCUS	AZB42795 469 bp DNA linear GSS 20-FEB-2001
DEFINITION	2M014I119F Mouse 10kb plasmid UUGCLM library Mus-musculus genomic clone UUGC2M014I119 F, DNA sequence.
ACCESSION	AZB42795
VERSION	AZB42795.1
KEYWORDS	GSS:
SOURCE	house mouse.
ORGANISM	Mus musculus
EUKARYOTA	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Neenen,E., Pedersen,T., Reilly M., Rose,M., Rose,R., Stokes,R., Tingey,G., von Niederhausern,A., and Wright,D., Weiss,R. Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts Unpublished (2000)
TITLE	Contact: Robert B. Weiss
JOURNAL	University of Utah Genome Center
COMMENT	University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0141 row: I column: 19 Seq primer: CGTTGTAACGACGCGGCAGT Class: plasmid ends High quality sequence stop: 469. Location/Qualifiers 1. .469 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="UUGC2M014I119" /clone_lib="Mouse 10kb plasmid UUGCLM library" /sex="Male" /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-" /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative
FEATURES	
source	

of pWD42 (gil14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 122 a 151 c 87 g 109 t
ORIGIN

Query Match 1.2%; Score 35; DB 17; Length 469;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1942 GTCGTGTCACCAAGTCATCAAGCTGGCTG 1976
|||||
Db 327 GTCGTGTCACCAAGTCATCAAGCTGGCTG 293

RESULT 6
BE553572 522 bp mRNA linear EST 15-AUG-2000
LOCUS ur47b04.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153391.5
DEFINITION similar to TR:054768 054768 A DISINTEGRIN AND METALLOPROTEINASE WITH THROMBOSPONDIN MOTIFS ; mRNA sequence.

ACCESSION BE553572
VERSION BE553572.1 GI:9817989
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 522)
REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1056147

Seq primer: -40RP from Gibco
High quality sequence stop: 385.
Location/Qualifiers
1. 522
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:3153391"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 116 a 133 c 158 g 115 t
ORIGIN

Query Match 1.2%; Score 35; DB 10; Length 522;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1942 GTCGTGTCACCAAGTCATCAAGCTGGCTG 1976
|||||
Db 73 GTCGTGTCACCAAGTCATCAAGCTGGCTG 107

RESULT 7
BF320986 305 bp mRNA linear EST 29-DEC-2000
LOCUS

DEFINITION uz57h10.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3673219.5
ACCESSION BF320986
VERSION BF320986.1 GI:11270026
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 305)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml

MGI:1433987
High quality sequence stop: 192.
Location/Qualifiers
1. 305
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3673219"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 79 a 71 c 85 g 69 t
ORIGIN

Query Match 1.1%; Score 30; DB 12; Length 305;
Best Local Similarity 100.0%; Pred. No. 0.0048;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1765 GGCTACCAACACAGCACCACCGGCTCACT 1794
|||||
Db 48 GGCTACCAACACAGCACCACCGGCTCACT 77

RESULT 8
AQ440250 400 bp DNA linear GSS 31-MAR-1999
LOCUS HS 5071.B2.D07.T7A RPCI-11 Human Male BAC Library Hmo sapiens
DEFINITION genomic clone Plate-647 Col-14 Row-H, DNA sequence.

ACCESSION AQ440250
VERSION AQ440250.1 GI:4551589
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 400)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (<http://info@resgen.com>). BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 647 row: H column: 14
Seq primer: T7
Class: BAC ends
High quality sequence stop: 400.

FEATURES
source

Location/Qualifiers
1. .400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=647 Col=14 Row=H"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

BASE COUNT
ORIGIN

100 a 116 c 98 g 83 t 3 others

Query Match

Best Local Similarity 1.0%; Score 29; DB 17; Length 400;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2164 GATGCAACTACTCGCTCTGAAGACAG 2192

|||||
DB 359 GATGCAACTACTCGCTCTGAAGACAG 387

RESULT 9
AW485848
LOCUS

68315 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence. EST 09-JUL-2000

AW485848

VERSION AW485848.1 GI:7055954

KEYWORDS EST.

SOURCE pig.

ORGANISM

Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE

AUTHORS

1 (bases 1 to 415)
Fahrenkrug, S.C., Fraking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W., and Keeler, J.W.

Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

Unpublished (2000)

JOURNAL

COMMENT

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with the phred v0.980904.e. Vector identified by cross_match with the -minscore 20 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCAGTCACGACG

Plate: 28 row: D column: 18

Seq primer: ATTTAGGTGACACTATAG.

FEATURES
source

Location/Qualifiers
1. .415
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

BASE COUNT 70 a 141 c 141 g 63 t

ORIGIN

Query Match 1.0%; Score 28; DB 10; Length 415;

Best Local Similarity 100.0%; Pred. No. 0.047;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2808 CCGCAAGCCCCAGGAGCTGGACTTCTGC 2835

|||||
DB 35 CCGCAAGCCCCAGGAGCTGGACTTCTGC 62

RESULT 10

B01986
LOCUS

DEFINITION B01986 169 bp DNA linear GSS 13-JUL-1996
CSRL-145A3-u CSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone CSRL-145A3, DNA sequence.

ACCESSION

B01986

VERSION B01986.1 GI:14111264

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 169)
Evans, G.A., Burbee, D., Davies, C., Hahner, L., Oliver, T., Gilbert, M., Jones, D., Ward, T., Gillilan, E., Schagemann, J., Probst, S., Harris, J., DeFord, J., McFarland, J., Burzinski, K., Khan, M., Kupfer, K. and Garner, H.R.

Genomic Sequence Sampled Map of Chromosome 11

TITLE

JOURNAL

COMMENT

Contact: Evans GA, Shane Probst
Mcdermott Center for Human Growth and Development
University of Texas Southwestern Medical Center At Dallas
5323 Harry Hines Blvd, Dallas TX 75235-8591
Tel: 214-648-1600

Fax: 214-648-1666

Email: gevanseutsw.swmed.edu, shane@mcdermott.swmed.edu

Seq primer: T7

Class: cosmid ends

High quality sequence stop: 169.

FEATURES

source

Location/Qualifiers
1. .169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSRL-145A3"
/clone_lib="CSRL flow sorted Chromosome 11 specific cosmid"
/sex="female"
/cell_type="chimeric hamster somatic cell hybrid"
/note="Vector: sCos-1; Human Chromosome 11 specific cosmid library prepared from flow sorted human Chromosome 11 derived from Chinese Hamster Ovary (CHO) monochromosomal somatic cell hybrid, J1"

BASE COUNT 26 a 56 c 41 g 36 t 10 others

ORIGIN

Query Match 0.9%; Score 27; DB 17; Length 169;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 475 CGCCGGGGTTCGGGGGGGGCTTCC 501

|||||

```

Db      56  CGCGGGGTGTCGGGGGGGCGCTTC 82
RESULT 11
BE281680
LOCUS
DEFINITION
601099657F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3491991 5',
mRNA sequence.
ACCESSION
BE281680
VERSION
BE281680.1 GI:9156727
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 627)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8537 row: a column: 16
High quality sequence stop: 530.
FEATURES
Location/Qualifiers
1..627
/organism="Mus musculus"
/strain="C2ECH II"
/db_xref="taxon:10090"
/clone="IMAGE:3491991"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT
153 a 159 c 177 g 138 t
ORIGIN
Query Match 0.9%; Score 26; DB 10; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1942 GTCTGTGTCGAAGGCAAGTCATCAA 1967
|||||
Db 95 GTCGTGTCGAAGGCAAGTCATCAA 120
|||||
RESULT 12
BB604671
LOCUS
DEFINITION
BB604671 RIKEN full-length enriched, 0 day neonate lung Mus
musculus cDNA clone E030005K13 5', mRNA sequence.
ACCESSION
BB604671
VERSION
BB604671.1 GI:11556073
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 247)
Aizawa,K., Akahira,S., Akimura,T., Araki,A., Arakawa,T., Carninci,P.,
Hanagaki,T., Hayatsu,N., Hirooka,T., Hirozane,T., Hodozawa,Y.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno
H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermolabile and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuina,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
Y., and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1..247
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E030005K13"
/clone_lib="RIKEN full-length enriched, 0 day neonate
lung"
/tissue_type="lung"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/notes="Site:1: SalI; Site:2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCGAGTTAATAATATCCCTCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT
47 a 30 c 128 g 42 t
ORIGIN
Query Match 0.9%; Score 25; DB 10; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1700 TGGAGCGCTGCCCCAGCTCAGCCTC 1724
|||||
Db 34 TGGAGCGCTGCCCCAGCTCAGCCTC 58
|||||
RESULT 13
BI03177
LOCUS
DEFINITION
BI03177 849 bp mRNA linear EST 26-JUN-2001
602889417F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5044493
5', mRNA sequence.

```

```

ACCESSION BI103177
VERSION BI103177.1 GI:14554070
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1122 row: a column: 06
High quality sequence stop: 743.

FEATURES
Source
Location/Qualifiers
1..849
/organism="Mus musculus"
/db_xref="FVB/N"
/clone="IMAGE:5044493"
/clone_lib="NCI CGAP Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library. |"
BASE COUNT 193 a 250 c 250 g 156 t
ORIGIN

Query Match 0.9%; Score 25; DB 13; Length 849;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1700 TGGAGCCCTGCCCGAGCTCAGCTC 1724
|||||
Db 540 TGGAGCCCTGCCCGAGCTCAGCTC 564

RESULT 14
BE666088 464 bp mRNA linear EST 25-APR-2001
LOCUS
DEFINITION Bos taurus
ACCESSION BE666088
VERSION 149119 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 464)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrnkug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keeler, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCATCACCAC
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 62 row: B column: 2
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..464
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT 73 a 157 c 163 g 71 t
ORIGIN

Query Match 0.8%; Score 24; DB 10; Length 464;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2812 AAGCCCCAGGAGCTGGACTTCTGC 2835
|||||
Db 215 AAGCCCCAGGAGCTGGACTTCTGC 238

RESULT 15
BI692986 664 bp mRNA linear EST 18-SEP-2001
LOCUS
DEFINITION 60334521F1 NCI CGAP_Mam2 Mus musculus cDNA clone IMAGE:5372261 5',
mRNA sequence.
ACCESSION BI692986
VERSION BI692986.1 GI:15655615
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 664)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11947 row: j column: 06
High quality sequence stop: 664.

FEATURES
Source
Location/Qualifiers
1..664
/organism="Mus musculus"
/db_xref="FVB/N-3"
/lab_xref="taxon:10090"
/clone="IMAGE:5372261"
/clone_lib="NCI CGAP Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 133 a 168 c 207 g 156 t

```

ORIGIN

Query Match 0.8%; Score 24; DB 13; Length 664;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2752 CTCACGTGTGTGGCCACGGAGGC 2775
|||||
Db 46 CTCACGTGTGTGGCCACGGAGGC 69

Search completed: May 2, 2003, 04:28:09
Job time : 3943 secs


```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-2

Query Match
  37.1%; Score 1917; DB 4; Length 837;
Best Local Similarity 46.3%; Pred. No. 2.6e-151;
Matches 381; Conservative 123; Mismatches 267; Indels 52; Gaps 16;

QY 1 MLLGLITLAFAGTAGGPEPEVVPVIRLDPDINGRRYYWRGPEDSGDOGLFQITAF 60
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 37 LLLLLASLLPSARLASLPREEIVFEKLSVL-----PGSGAPARLLCLRQAF 88
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 61 QEDFYLHTPDQAFTAPASTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDSEAAVSLC 120
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 89 GETLLELEQDSGVQVEGLTVQYLQAPE-LLGGAEP--GYLICTINGDPESVASLHWD 145
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 121 GG-LRGATGYRGAEEVVISPLPNASAPAAQRNSQGHLLQRGVGGPGSDPTSRGCVASG 179
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 146 GALLGVLYQRAELHQLPLEGGTPNSA--GGPCAHLIRK-----SPASGQGMCNV--- 196
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 180 WNPAILRALDPYKPRAGGFSRSGRRGRKRFVPIRYVETLVVADESVMKEFGADLE 239
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 197 -----KAPLGSPSPRP--RAKRFASLSRFVETLVVADKMAAFHAGLK 239
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 240 HYLLTLLATAARLYRHPISILNPINIVVYVLLLRDRSGPKVTGNAALTILNFCAWOKKL 299
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 240 RYLLTVMAAAAKAFKHSIRNPVSLVTRVLILSGEGEPQVGSAAQTLRSFCAWQRL 299
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 300 NKVSDKHEPYWDATLFTRODLGCATCTDILGMADVGTWCDPKRSCSVIEDDGLPSAFTT 359
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 300 NTPEDSDPDHFDATLFTRODLGCATCTDILGMADVGTWCDPARSCAIVEDDGLQSAFTA 359
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 360 AHELGHVFNMPHDNVKVEEYFVKL--RANHMSPTLIQIDRANPWSACSAAIITDFLDSG 418
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 360 AHELGHVFNMLHNSKPCISLNGPLSTSRHYMAPVMAHVDEEPFWSCSRFTIDFLDNG 419
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 419 HCDCLLDQSPKIPSLPELPCASTYLSQCELAFGVSKGPCPYM-QYCTKLWCTGKAKGQ 477
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 420 YGHCLLDKPEAPLHLPVTFPGKDYDADRQCOLTFGPDSPRHCPQLPPPCAALWCSGHLNGH 479
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 478 MYCQTRHPWADGTSCGBGKLCACVERHNLNKHVD--GSWAKWDYPGPCSRTCGG 535
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 480 AMCQTHSPWADGTGCPGAQACMGRCGLHMDQLQDFNIPQAGGPGWPGWDCSRTCGG 539
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 536 VOLARRQCTNTPANGGKYCEGVRYKRSCLNLEPCPSASGKSPREBQCEAFNGYNHSTN 595
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 540 VQFSRDCTRVPRNNGKYCEGRTFRSCNTEDCP--TGSALTFREQCAA---YNHRTD 595
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 596 RLTL---AVAMVPKYSGVSPDKCKLICRANGTYFVVLAPKVVVDGTLCPDSTSVQCQG 652
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 596 LFKSFGPMQWPRYTGVAPQDQCKLTQARALGYVYVLEPRVVDGTPCSPDSSVVCVQG 655
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 653 KCIKAGDNLGSKKRDCKVCGGDNKSKCTGLTCKPMHGVNFVVAIPAGASSIDIR 712
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 656 RCIHAGCDRIIGSKKFDKCMVCGGDSGSGSKSGSRFRYGYNNVVTIPAGATHFLVR 715
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 713 QRGYKGLIGDNNYLALNKSQKYLNLNGHFVYSAVERDLVVKGSL-LRYSCTGTAVESLQA 771
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 716 QQGNPG--HRSIYLALKLPDGSYALNGEYTLMPSPDVLPLGAVSLRYSGATASETLSG 773
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

QY 772 SRPILPLTVLTVSVGKMTPPRYKSYFLPKPEPREDKSSHPKD 814
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 774 HGPLAQPLTLQVLVAGNPQDTRLRYSEFFVPRPTPTPTPDQ 816
   :|||: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

RESULT 4
US-09-130-491-13
; Sequence 13, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
```

Sequence 2, Application US/09369364A
Patent No. 6391610
GENERAL INFORMATION:
APPLICANT: Apte, Suneel
APPLICANT: Hurskainen, Tiina L.
TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
FILE REFERENCE: 26473/4007/10-30-00
CURRENT APPLICATION NUMBER: US/09/369,364A
CURRENT FILING DATE: 1999-08-06
- NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 930
TYPE: PRT
ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

Query Match 35.9%; Score 1854; DB 4; Length 930;
Best Local Similarity 43.1%; Pred. No. 5.4e-146;
Matches 369; Conservative 116; Mismatches 265; Indels 106; Gaps 13;

QY 90 GLTGSSDLRCFCYSGDVNAPDFAAVSLCGGLRGAGFYGAEXVISPLPNASAPAAOR 149
DB 121 GLSASSHGRHCFYGTGDSFRLAVFDLCGLDGGFAVKHARYTLKPLRGSWAERYER 180
QY 150 -----NSOGALLORRG-----VPGGSPGDPTSRGCVGASGNPAILRALDPY 191
DB 181 TYDGSRRILHYNREGSFALPRASCETPASP----- 217
QY 192 KPRRAGFGESESRSSG-----RAKRFVSIPIRYVTLVAD 227
DB 218 -PQESPVSRSRRSALAPQLLDHSAFSPGNAGPQTWRRRRRSISRARQVELLIVAD 276
QY 228 ESMVFGADLEHLLTAAARLYRHSILNPINIVVVKVLLLRDSDPKVTGNAAL 287
DB 277 SSMARYGRGLOHLYLTTWASIANRLYSHASENHRLAVKVVLTDKDTSLEVSNAAT 336
QY 288 TLNRCFAWKLNKYSDKHPEYWDTAILFTRODLCGATCTDLGMADVGTMDPKRSCSV 347
DB 337 TLKNECKQHOHNLGDDHEHYDAAILFTREDLCGHSHCDTLGMADVGTICSPERSCAV 396
QY 348 IEDDGLPSAFTAHGLHGVNPHDNVNVCEVFGKLRANHMSPTLIQIDRANPWSACS 407
DB 397 IEDDGLHAFTVAHEIGHLGLSHDDSKFCEENFTEDKRLMSSILTSIDASKPWSKCT 456
QY 408 AAIITDFLDGSHGDCLLDQSPKIPSLPDLPGASYTLSSQCELAFAFGVSKPCPYMYCTK 467
DB 457 SATITEFLDDHGNCLLDLPRKIILGPPELPGQTYDATQOCNLTFGPEYSVCPGMDVCAR 516
QY 468 LMCTGKAGQWYQTRHFPWADGTSCEGSKLCLGACV-----HNLKRVDSWAKW 522
DB 517 LWCVVROGMVCLTKLPVAVESTPCGGRVCLQCKVDKTKKYISTSSH---GNMGSW 573
QY 523 DYPGCSRTCGGVOLARQCTNPYPANGKYCEGVRYKRSNLEPCPSASGKSFREE 582
DB 574 GPWGQSCSGGVQFAYHRCNNPAPRSGRYCTCKRAIYRSCSVTPCP--PNGK\$FRHE 631
QY 583 QCEAFNGYHNSHNLTLAVAWVPKYSGVSPRDKCLICRANGTYFYVYLPKVDGTICS 642
DB 632 QCEAKNGYQSDAKGVKTFEWWPKYAGVLPADVCKLTCAKGTGYVYVSPKVTDTGECR 691
QY 643 PSTSVCVGCKIKACDGNLGSKRFDKCGVCGDNKSCCKVTGLTKPMHGVNFVAI 702
DB 692 PYSNSVVRGRCVTRGCDGIIGSKLOYDKCGVCGDGNDSCTKIITFNKSKGYTDVRI 751
QY 703 PAGASSIDIRQRYKGLIGDDNYLAKNSQGRYLLNGHFVSAVERDLVWGLSLRYSGT 762
DB 752 PEGATHIKVYQKAKDQTFEPAYLALKKKTGEYLINGYMTSETIIDINGTVNMYSGW 811
QY 763 GRAVESL-----QASRPILPELTVVLSVGKMTPPRVKYSFYLKPEPREDSK---SHPKD 814
DB 812 SHRDDPLHGMGYSATKEI---LIVQILATDPTKALGVRYSEFPVPRKTTQKVNVSIVSHGSN 868

QY 815 PRGFSVLHNSVLSLSNQVBPDDPPARVWAGSWGPCSSACSGLOKRAVDCRGSAQORT 874
DB 869 KVGPP---HSTQL-----QWVTGFWLACSRCTCDTGWHTRTVQCQ-DGNRKL 909
QY 875 VPACDAARHPVETOAC 890
DB 910 AKGCLLSQRPSAFKOC 925

RESULT 6
US-09-369-364A-13
Sequence 13, Application US/09369364A
Patent No. 6391610
GENERAL INFORMATION:
APPLICANT: Apte, Suneel
APPLICANT: Hurskainen, Tiina L.
TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
FILE REFERENCE: 26473/4007/10-30-00
CURRENT APPLICATION NUMBER: US/09/369,364A
CURRENT FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 1882
TYPE: PRT
ORGANISM: Homo sapiens ADAMTS-9
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (468)
OTHER INFORMATION: Xaa = C
NAME/KEY: MOD_RES
LOCATION: (521)
OTHER INFORMATION: Xaa = Y
US-09-369-364A-13

Query Match 35.8%; Score 1848.5; DB 4; Length 1882;
Best Local Similarity 39.6%; Pred. No. 4.4e-145;
Matches 396; Conservative 144; Mismatches 340; Indels 121; Gaps 23;

QY 22 EREVVPVIRLDP-----DINGRRYVW-----RGPEDSGQGLIFQITAFQ 61
DB 49 EYEIVSPIRYNALGEPPTNVHFKRTRRSINSATDPAPAFASSSSSTSSQAHYRLSAFG 108
QY 62 EDFYHLTPDQAFLAPAFSTEHLGP-----LQGLTGSSDLRCFCYSGDVNAPDFAAV 117
DB 109 QQFLNLTANAGFIAPLFTVTLGTPGVNQTKEYSEEAELKHCFYK----- 155
QY 118 SLGGLRGAGFYGAEXVISPLPNASAPAAORNSOGAHLIQRGVPGSPGSDPTSRGVA 177
DB 156 RLQOYQLRANG-----RHQPLLRNEKHNHSDKKKTRARKWGERINLAGDV 202
QY 178 SGWNPAI-LRALDPYKPRRAGFGESESRSSRAKRFVSIPIRYVETLVVVADESQWVFGA 236
DB 203 AALNSGLATEAFSAYGNKTDNTRKTRHRT---KRFLSYPREVEVLVADNRMVSYHGE 259
QY 237 DLEHYLLTLTAARLYRHSILNPINIVVVKVLLLRDSDGPKVTGNAALTILRNFCAWQ 296
DB 260 NLQHYLLTMSIVASIIYKDFSIGNLINIVLVLIHNEQDGSISFNAQTILKNFCOMQ 319
QY 297 KKLNVKSDKHPEYWDTAILFTRODLCGA-TTCDTLGMADVGTMDCKRSCSVTIEDGLPS 355
DB 320 HSNSPGGIHH-----DTAVLLTRQDICRAHKCDTLGLAELGTICDPRYRSCSISEDSGLST 375
QY 356 APTAHGLHGVNPHDNVNVCEVFGKLRANHMSPTLIQIDRANPWSACSAAITDFL 415
DB 376 APTAHGLHGVNPHDNVNVCEVFGKLRANHMSPTLIQIDRANPWSACSAAITDFL 434
QY 416 DSHGDCLLDOP-SKPISLPEDLPASVYTLSSQCELAFAFGVSGKPCPYMYCTKLWC---T 471
DB 435 DTGEGECLLNEPESRPYLPQLPGLIYNVKNQKELIFGQSVQCPYMYCRRLLWCNNVN 494

Db	247	-AGGSGPQTWRRRRS-----ISRARQVELLLVADASMARLYGRLQHUYLLTLASIAN	299
Qy	251	RLYRHPSTLNPIVVVKVLLLRDROSGPKVTGNAAALTLRNFCAMQKLNKVSKDHPEYW	310
Db	300	RLYSHASINHLRAVKKVYVVLVDKSLSEVSKNAATLTKNFCKWHQHGLGDHBEHY	359
Qy	311	DYAILFTRODLGGATCDTGLGMADVPMCDPRSCSVIEDDGLPSAFTTAHELGHVENMP	370
Db	360	DAALFTREDLCGHISCDTLGMADVGTICSPERSCAVIDEDGLHAFTVAHEIGHLLGLS	419
Qy	371	HDNVKCEVFCKLRANHMSPPTLIQIDRANPWSACSAALITDFLDLSHGDCILLDQPSKP	430
Db	420	HDSKFCEETFGESTEDKLMSILTSIDASKPMKSCTSATITEFLDDGHGNCLLDLPRKQ	479
Qy	431	ISLPEDLPASYYTLSSQCELAFGVGSKPCPYMYCQTKLMCTGKAKQMVQCOTRHFPWADG	490
Db	480	ILGPPELPQGYDATQQCNLTFFGPEYSVCPGMDVCARLWCVVVRQGQVMCLTKKLPAVEG	539
Qy	491	TSCGEKCLKAGACVER-----HNLNKHRYDVGSWAWDYDGPCSTTCGGVGOLARRCOTN	545
Db	540	TPCGKGRIQLQKCVDKTKKYIYSTSSH----GNWGSWGSGOCSSRCGGGVQFAYRHNN	596
Qy	546	PTPANGKYCEGRVYKRVSCNLPCPPSSASGSKFREEOCEAPNGYNHSTNRILTAVAWVP	605
Db	597	PAPRNNRGYCTKRAIYRCSLMPCP--PNGKSFREHQCEAKNGYQSDAKGVKTFWEVWP	654
Qy	606	KYSGVSFRDKCLICRANGTYFYVVLAPKVVYDGTCLSDPSTSVCYOGKCIRAGCDGNLGS	665
Db	655	KYAGVLPADVCKLTCRANGTYVYVFSPKYTDTECRPYNSVCYRGKVRTGCDGLIGS	714
Qy	666	KRFDFKCGVCGDNKSKCKVTGLTKPMHHGNFWIPAAGASSIDI RORGYTKGLIGODNY	725
Db	715	KLOYDKCGVCGDNSCTKIYGTEHNKSKGYTDVARIPEGATHIKVROFKAKDOTRFTAY	774
Qy	726	LALKNSQGYKLLNGHFVYSAVERDLVKGSLRYSGTGTAVESL-----QASRPILPLEPLT	780
Db	775	LALKKNGEYLINGKYMISTSETTIIDINGTVMNYSCWSHRDDFLHMGYSATKEI---LI	831
Qy	781	VEVLSVGKMTPPRVRSYLEPKAPEBKDSHPKDPGRCPSYLHNSVLSL-SNOVQPDOPDRP	839
Db	832	VQILATDPTKPLDIVRYSFVFPKK-----STPK-----VNSTVSHGNSKGVSHTSQP	877
Qy	840	PARWAGSWGPCSASCGLQKRAVDCRGSAGORTVTPACDAHRPVETQAC	890
Db	878	--OWTGPWLACSRTCDTGTHTRVOCO-DGNRKLAQKCPCLSORZSAFKQOZ	925

[illegible]

```
QY    781   VEVLVSGKMTPRVRYSFYLEKEPREDKSHGHPDGRGPSVLHNSVLSL-SNOVEQPDDR      833
Db     832 VQLIATDPTKPLDLRVISFEVPKK-----STPK-----VNVSITSGSNKVGSHTSQP      877

QY    840 PARWAGSWGSPCASCSGLGKLQRVAIDCRSAGORTVTPACDAHRPVETOAC      890
Db     878 --QWVTGPWLACSRTCDDTGWHTRTVOQQ--DCNKRKLAKGCPLQSPPSAFKQC      925


RESULT 8
US-09-369-364A-15
; Sequence 15, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369, 364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 874
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-9
US-09-369-364A-15

Query Match          34.8%; Score 1798.5; DB 4; Length 874;
Best Local Similarity 42.18; Pred.No.2.le=141;
Matches 369; Conservative 122; Mismatches 302; Indels 83; Gaps

QY    115 AAYSLVCGLLRGFYGAEYYVISLPNASAPAAQRNSOQHALLRRRGGPGSDPTSRC      174
```

```
Query Match          34.8% ; Score 1798.5; DB 4; Length 874;  
Best Local Similarity    42.1%; Pred. No. 2.1e-141;  
Matches 369; Conservative 122; Mismatches 302; Indels   83; Gaps  
  
Qy      115 AAVSLCGLLRGAFYGYGAAYVISLPNAPAFQAQRNSQAHLQLRRGVPGSGDPISRRC 174
```

; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-130-491-16

Query Match 31.1%; Score 1607; DB 4; Length 551;
Best Local Similarity 48.5%; Pred. No. 1e-125;
Matches 282; Conservative 108; Mismatches 152; Indels 40; Gaps 10;

QY 377 CREVFGLKIRANHMSPITLIQIDRANPWSACSAAITDFLDSHGDCLLDOPSKPISLPEB 436
DB 1 CASLNGVSDSHLMSLSSLDHSPWSPCSAYMVTSLDNGHGECLEMDKPNKIPSPD 60
QY 437 LFGASTYLSQOCELAFGVGSKPCP-YMQYCTKLWCTGKAKGOMVQOTRHFPAWADTSCGE 495
DB 61 LPTGLDANROQOFTFEESKHCPCDAASTCTTLMCTGTSGLLVQCTKHFPWADTSCGE 120
QY 496 GKLCAGACVERHNKHKH--RVDGSWAKWDPYPCSRCTCGGVQVLARQCTNPTPANG 552
DB 121 GKWCVGSKVNTDM-KHEATPVHGSWGPWPGDCSRCTCGGVQVYTMRECONPVKNGG 179
QY 553 KYCEGVRYKRYSCNLEPCPSSASGSKSFREEQCEAFNGYHNSTNRLTLAVAVPKYSGVSP 612
DB 180 KYCEGRVRYSCNIEDPCDN-NGKTFREBQCEAHNEFSKASFGNEPTVEWTPKYAGVSP 238
QY 613 RDKCKLICRANGTYFVLAPKVVDCTLCSPDSTSVCVQKCIKAGDCNGLSKKRFDC 672
DB 239 KDRCKLTCEAKGIGYFVLQPKVVDGTPLCSPDSTSVCVQKQCVKAGCDRIIDSKKFKDC 298
QY 673 GVCGGNCKSKVTGLFTKPMHGYNFVAIPAGASSIDIRQYKGLIGDNDYALKNSQ 732
DB 299 GVCGGNSTCKKSGIVTSPCYHDIPTIPAGATNIEVKHRNQSRNNGSFLAIRAAD 358
QY 733 GYLLNGHFFVSAVERDLVVGSLRYSGTGTAVESLQASRPILPTVEVLSVGKMTTP 792
DB 359 GTYILNGNFTSLTEQDLYTKGTVLRYSGSAALERISFSPLEKPLTTQVLMVGHALRP 418
QY 793 RVRYSFYLPKEPREDKSSHPKDPGRPSVLNSVLSINQVEQPDPRPPARWAGVSGPCS 852
DB 419 KIKFTYFMKKTES-----FNAIPTFS-----EWVTEEGECS 451
QY 853 ASCSGSLQKRAVDCRSAGORTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCKSKC 909
DB 452 KTCGSGWQRVWQCRDINGH---PASECAKEVPKASTRCPADLPCPHWQVGDWSPCKSKC 508
QY 910 GRGFORSLKCVGHGGRLLARDQCNLHRKQPE-LDFCVLRPC 950
DB 509 GKGYKKRILKCVSHDGVLSNESCDPLKPKHYIDFCTLTQC 550

RESULT 10
US-09-130-491-8
; Sequence 8, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gooddearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 481

10

10

600

200 000

QY 188 LDPYKRRAGFSGSRSSRRSRAKRFVSIPIRYVETLVVADESVMKPHG-ADLEHYLLTLL 246
Db 268 PDEYK---SCLRHKRSLLSHRNEEL-----NVETLVVDDKMMQNHENITTVLITL 319
QY 247 ATAAIRYRHPISILNPINIVVVKVLLLRDRSDGPKVTGNAAALIRNFCAMQKKL-NKVSOK 305
Db 320 NMVSALEKDTIGNINIAVGLILLEDEQGLVISHHADHTLSSPCQWOSGLMGKDGTR 379
QY 306 HPEYWDTAIFTRODLGC--ATTCDILGMADVGTCDPKRSCSVIEDDGLPSAFTTAHEL 363
Db 380 H---DHAILLTGIDCSWKNRPECDTLGFAPISGMSKYRSCCTINEDTGLGLAFTTAHES 435
QY 364 GHVFNPHDNV-KVCERFVKRLRANHMSPTLIQIDRANPWSACSAAIITDFLDSHGDC 422
Db 436 GHNFMIHDGEGNCKKSEG-----NIMSTLAGRNCVFSWSPCSROYLHKFLSTAQAIC 490
QY 423 LLDQPSKPI---SLPEDLPAGSYTLQQCELAFAFGVSKPCPY---MOYCTKLMC---TGKA 474
Db 491 LADQP-KPVKEYKPEKLPCELYDANTQKWQFGEKAKLCMLDFKKDICKALWCHIRGRK 549
QY 475 KGWVCOTRHFPRADFTSCGEGKLCIKGACVERHNLNKHVRDGSMAKWDPYGPCSRTCGG 534
Db 550 -----CETKMPAAEGTICGHDWCRGQCVKYGDGPKPTHGHSWDSWSPCSRCTCGG 604
QY 535 GVQLARROCTNPTPANGKCEGVYRVYRSCNLEPCPSSASGKSFREEQC-----EAFNG 589
Db 605 GVSHRSLCTNPKPSHGKCEGSTRILKLCNSOKCPRDS--VDFRAAQCAEHSRRFRG 662
QY 590 YNHSNRLTLAVAWPKYSGVSPRDKCKLICRANGTGYFYVLAPKVVDGTLCSPDSTSV 649
Db 663 RHYK-----WKP-YTQVEDQDLCKLYCABEGDFEFLSKNVKVDGTPCSEDSRNV 712
QY 650 VOGKICRACGDNGLSKRDKCVCVCGDNKSKCKVTGLTKPMH--GYNFVVAIPAGAS 707
Db 713 IDGICERVCDNVGSDAIVEDVCGVCGNNSACTIHRGLTKHHHTNQYHMWITPSGAR 772
QY 708 SIDIRQRYKGLIGDNYLALNKSQGYLLNGHFVVSAAVERDLVVKGSLLRYSGTGTAVE 767
Db 773 SIRIVEMNV-----STSVISVRNALRYYLNGHWTVDWPGR-YKFSGTTFDYYRSYNEPE 826
QY 768 SIQASRPLEPLTVELSVGKMTTPRVYRYSYLPKEPREDKSSHPKDPGRPS-----819
Db 827 NLIATGPTNETLIVELLEFGGR--NPGVAEYSMPR-----LGTEKQPPAQPSYTWAIYRS 879
QY 820 -----VLHNSVLSLSNOVE-----QPDDR-----PARKVAGSNGPC 851
Db 880 ECSVSCGGQMTVRGECYRDLKFQVNSFCNPKTRPTVTLVPCVKSACPPSSVGNNSAC 939
QY 852 SASCGSGLOKRAVDC--RGSAGQRTVPA--CDAARHPVETOAC-GEPC-PTWELSANSPC 905
Db 940 SRTCGGAQSRPVQCTRRVHYDSEVPASLCPQP-APSSRQACNSQSCPPAWSAGPMAEC 998
QY 906 SKSCGRGQRRSLKCVGHG---GRLARDOCNLHRKPOELDVCFLRPC 950
Db 999 SHTCGKGRKRAVACKSTNPSARAOLLDPDVACTSEPKPRMHEACLLQRC 1047

RESULT 13
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Procop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE.
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York

STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/491.522
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886.333
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8389-0060-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-491-522-5

Query Match 23.6%; Score 1217; DB 4; Length 1211;
Best Local Similarity 33.1%; Pred. No. 1.1e-92;
Matches 347; Conservative 133; Mismatches 405; Indels 162; Gaps 47;
QY 9 LAFAGRTAG---GFEPEREVVVPIRLDPD-----INGR-----YYW 42
Db 39 LAAADPPGGLPGHGAERILAVPVRTDAQRLVSHVSAATSRAQVARRAAVPTSPFP 98
QY 43 RGEDSGDGLFIQITAFQEDFYHLTTPAQFLAPAFSTEHLGVPLQGLTGSSDLRRCF 102
Db 99 GNEEEFGSHLFYVNVFGRDLHLRLRPNARLVAPGATMEWGE--KGTRVEPLLGSL 156
QY 103 YSGDVN--AEPDFAAVSLCGGLRGAFYRGAEYVISPFPNAPASAAQNRSG-AHLQ 159
Db 157 YGVDVAGLAFASS-VALSNCGLAGLIRMEEEFFTEPLEKGL--AAQAEQGRVHVYR 213
QY 160 RGVPGGPGSDPTSRGVASGNPAILRALDPYKPRAGFGESRRSRRAKRFVSIPIY 219
Db 214 RPPTSPPLGGPQA---LDTG---ASLDSLSLS--RALGVLEEHSRRRRRAHADDD 266
QY 220 VETLVVADESVMKPHGAD-LEHYLLTLLATAARLYRHPISILNPINIVVVKVLLLRDRS 277
Db 267 NEIVLLGVDDSVVQFHGKEHVQKYLITLMNVNEIYHDESAGHINVLIRIILSYGKS 326
QY 278 GPKV-TGNAALTNRNFCAMQKKLNKYSKHPYWDTAIFTRODLGCATTCTDGLMADV 336
Db 327 MSLEIGNQSQSLNVCRWAYLQPKPTDGHDEYDHAIFLTRQDF-GPSGMQ--GYAPVT 383
QY 337 TMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCERFVKRLRANHMSPTLI 395
Db 384 GMCHPVRCTLNHEDGFSSAFVVAHETGHVLMGHDDGQGNRCGD---EVLGLSAPLQV 440
QY 396 QIDRANPWSACSAAIITDFLDSHGDCCLDOP--SKPISLPEDLPASVTLSSQCELA 452
Db 441 AAFERFHSRCSQOELSRYLHS--YDCLLDDPPFAHDWP-ALPQ-LFGLHYSMNEQCRFD 496
QY 453 GVSGKPCPYQWY---CTKLMCTGKAKGQWVCQTRHPEWADGTSCGEGKLCIKGACV----505
Db 497 GLGYMCTAFRTFDPCKQLWCS-HPDNPYFCKTKKGGPLDGTMCAPGKHCFGKICWLT 555
QY 506 ---ERHNLNKHVRDGSMAKWDPYGPCSRTCGGVQVLAARROCTNPTPANGKCEGVYRV 562

[illegible]

```

RESULT 15
US-09-369-364A-17
; Sequence 17, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-10

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:13:53 ; Search time 49 Seconds
(without alignments)
2583.430 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 950
Sequence: 1 MLLGLTLAPAGRTAGGFE.....DQCNLHRKPOELDFCVLRPC 950

Scoring table: OLIGO

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	950	100.0	950	22	AA62299 Human metalloprotease
2	931	98.0	950	23	AAE22541 Human metalloprotease #2
3	673	70.8	952	23	AAU74751 Human protease PPT
4	635	66.8	928	23	AAU72899 Human metalloprotease
5	345	36.3	367	23	AAE22542 Human metalloprotease #3
6	300	31.6	321	23	AAE22540 Human metalloprotease #1
7	177	18.6	381	21	AAE21261 Human metalloprotease
8	154	16.2	505	21	AAE21257 Rat metalloprotease
9	17	1.8	727	20	AAW78435 Human ADAMTS-1 pro
10	17	1.8	896	21	AAE21265 Mouse metalloprotease

11	17	1.8	950	20	AA49501 Human METH1 protei
12	17	1.8	950	21	AA53899 Amino acid sequenc
13	17	1.8	950	22	AA73549 Human ADAM-type me
14	17	1.8	950	22	AA50002 Human METH1 - Homo
15	17	1.8	967	19	AAW80285 Human integrin lig
16	17	1.8	967	20	AAW04142 Human tango-71 pro
17	17	1.8	967	20	AAW78189 Human secreted pro
18	17	1.8	968	22	AA50011 Protein; SEQ ID 12
19	14	1.5	180	22	AA55532 Human protein sequ
20	13	1.4	107	22	AA004896 Human polypeptide
21	13	1.4	625	23	AAW48394 Rat aggrecanase
22	13	1.4	837	20	AAW75425 Human aggrecan deg
23	13	1.4	837	21	AAW9429 Human PRO1563 (UNQ
24	13	1.4	837	22	AAW78228 Human aggrecanase-
25	13	1.4	837	22	AAU29199 Human PRO polypept
26	13	1.4	840	21	AA66178 Human metalloprote
27	13	1.4	874	22	AA72287 Murine ADAMTS-9 am
28	13	1.4	947	22	AA86950 Human metalloprote
29	13	1.4	958	21	AAE21255 Human metalloprote
30	13	1.4	1073	21	AAE21264 Human metalloprote
31	13	1.4	1602	23	ABG30702 Human aggrecanase
32	13	1.4	1629	23	ABG30703 Human aggrecanase
33	13	1.4	1629	23	AAO14448 Human ADAMTS-SI pr
34	13	1.4	1882	22	AA72286 Human ADAMTS-9 am
35	13	1.4	1907	23	AAU77133 Human metalloprote
36	13	1.4	1916	23	AAE19173 Human metalloprote
37	13	1.4	1934	22	AA72286 Human metalloprote
38	13	1.4	1935	23	AAU72896 Human metalloprote
39	13	1.4	1935	23	AAU72896 Human metalloprote
40	12	1.3	12	21	AAE21277 Human ADAMTS 1 Zn
41	12	1.3	12	21	AAE21262 Human ADAMTS 9 Zn
42	12	1.3	98	21	AAE21262 Rat metalloprote
43	12	1.3	481	20	AAW04145 Rat Tango-76 prote
44	12	1.3	905	22	AAW72284 Murine ADAMTS-8 am
45	11	1.2	58	22	AAW90625 Human secreted pro

ALIGNMENTS

RESULT 1
AA62299
ID AAG62299 standard; protein; 950 AA.

XX AAG62299;

AC AC

XX 23-AUG-2001 (first entry)

DT Human metalloprotease MDT56 protein.

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

XX osteopathic; antiarthritic.

XX Homo sapiens.

XX WO200134785-A1.

XX 17-MAY-2001.

XX 10-NOV-2000; 2000WO-JP07917.

XX 11-NOV-1999; 99JP-0321740.

XX 16-MAY-2000; 2000JP-0144020.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

XX (KAZU-) KAZUSA DNA RES INST.

XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI: 2001-343602/36.

XX DR N-PSDB; AA41003.

XX Metalloprotease with aggrecanase activity for treating joint diseases

PT especially osteoarthritis -
XX Claim 1; Page 56-60; 85pp; Japanese.
XX This invention relates to a metalloprotease with aggreganase activity.
CC The invention includes protein and DNA sequences of the metalloprotease,
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC the metalloprotease of the invention termed MDT56.
XX
SQ Sequence 950 AA;

Query Match 100.0%; Score 950; DB 22; Length 950;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 950; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDSDGGLIFQITAF 60
Dy 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDSDGGLIFQITAF 60
Qy 61 QEDFYHLTPDAQFAPAFSTHGLVPLQGLTGSSDLRRCFCYSGDVNAEPDSFAAVSLC 120
Dy 61 QEDFYHLTPDAQFAPAFSTHGLVPLQGLTGSSDLRRCFCYSGDVNAEPDSFAAVSLC 120
Qy 121 GGLRGAFYGAERVIISPLNASAPAOQNSOGAHLQRRGPGSPGSDPTSCGVASGW 180
Dy 121 GGLRGAFYGAERVIISPLNASAPAOQNSOGAHLQRRGPGSPGSDPTSCGVASGW 180
Qy 181 NPAILRALDPKPRRAGFSGRSRRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 240
Dy 181 NPAILRALDPKPRRAGFSGRSRRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 240
Qy 241 YLLTLLATAARLYRHPSTLNPINIVVVKVLLLRDRDSGPKVTGNAALTLLNFCAWQKLN 300
Dy 241 YLLTLLATAARLYRHPSTLNPINIVVVKVLLLRDRDSGPKVTGNAALTLLNFCAWQKLN 300
Qy 301 KVSQKHPEYWDATLFTRODLGATTCDTLGMADVGMCDPKRSCSVIEDGLPSAFTTA 360
Dy 301 KVSQKHPEYWDATLFTRODLGATTCDTLGMADVGMCDPKRSCSVIEDGLPSAFTTA 360
Qy 361 HELGHVFNHNDNVKVEEVEFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDSGHG 420
Dy 361 HELGHVFNHNDNVKVEEVEFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDSGHG 420
Qy 421 DCLDQPSKPISLPDLPGASYTLLSQCELAFGVGSKPCPYMQYCTKLWCTGKAKGQWVC 480
Dy 421 DCLDQPSKPISLPDLPGASYTLLSQCELAFGVGSKPCPYMQYCTKLWCTGKAKGQWVC 480
Qy 481 QTRHPPWADGTSCGEGKLCIKAGACVERHNKLRVDSWAKWDYGPSCRTCGGVQVQLAR 540
Dy 481 QTRHPPWADGTSCGEGKLCIKAGACVERHNKLRVDSWAKWDYGPSCRTCGGVQVQLAR 540
Qy 541 RQCTNPTPANGKYCEGVYRVYRSCNLEPCPSASGKSFREEOCEAFNGYHNHSTNRLTA 600
Dy 541 RQCTNPTPANGKYCEGVYRVYRSCNLEPCPSASGKSFREEOCEAFNGYHNHSTNRLTA 600
Qy 601 VAWVPKYSGVSPROCKKILCRANGTGTFYVLAPKVDGTLCSPOSTSVCGQKCIKAGCD 660
Dy 601 VAWVPKYSGVSPROCKKILCRANGTGTFYVLAPKVDGTLCSPOSTSVCGQKCIKAGCD 660
Qy 661 GNLGSKKRFDCGCGGDKNSCKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQRYKGLI 720
Dy 661 GNLGSKKRFDCGCGGDKNSCKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQRYKGLI 720
Qy 721 GDDNYLALKNQSGKYLKNGHFVWSAVERDLVVKSLRYSGTGTAVESLQASRPILPLEPT 780
Dy 721 GDDNYLALKNQSGKYLKNGHFVWSAVERDLVVKSLRYSGTGTAVESLQASRPILPLEPT 780
Qy 781 VEVLVSGKMTTPRVRYSYFLPKPEDEKSSHPKDRGSPVLSLNSNOVEQDDRRPP 840
Dy 781 VEVLVSGKMTTPRVRYSYFLPKPEDEKSSHPKDRGSPVLSLNSNOVEQDDRRPP 840

Db 781 VEVLVSGKMTTPRVRYSYFLPKPEDEKSSHPKDRGSPVLSLNSNOVEQDDRRPP 840
Qy 841 ARWVAGSWGSCSASGSLQKRAVDRCGSAQRTVPACDAARHPVETQACGECPTWELS 900
Dy 841 ARWVAGSWGSCSASGSLQKRAVDRCGSAQRTVPACDAARHPVETQACGECPTWELS 900
Qy 901 AWSPCSKSGRGFQRRSLKCVGHGGRLLARQCNLHRKPQELDFCVLRPC 950
Dy 901 AWSPCSKSGRGFQRRSLKCVGHGGRLLARQCNLHRKPQELDFCVLRPC 950

RESULT 2
AAE22541
ID AAE22541 standard; Protein; 950 AA.
XX
AC AAE22541;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #2.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
OS Homo sapiens.
XX
PN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Friddle CJ, Hilbun E;
DR WPI; 2002-372123/40.
DR N-PSDB; AAD35569.
XX
PT Novel nucleic acid encoding a human protease, useful as a hybridization
PS probe for screening libraries and assessing gene expression patterns -
PS Claim 6; Page 36-38; 41pp; English.
XX
CC The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridisation probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 950 AA;

Query Match 98.0%; Score 931; DB 23; Length 950;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 931; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 EPEREVVVPIRLDPDINGRRYYWRGPDSDGGLIFQITAFQEDFYHLTPDAQFAPAF 79
Dy 20 EPEREVVVPIRLDPDINGRRYYWRGPDSDGGLIFQITAFQEDFYHLTPDAQFAPAF 79
Qy 80 STEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDSFAAVSLCGLRGAFYGAERYVISPL 139
Dy 80 STEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDSFAAVSLCGLRGAFYGAERYVISPL 139
Qy 140 PNASAPAOQNSOGAHLQRRGPGSPGSDPTSCGVASGNPAILRALDPKPRRAGFG 199
Dy 140 PNASAPAOQNSOGAHLQRRGPGSPGSDPTSCGVASGNPAILRALDPKPRRAGFG 199

Db 140 PNASAPAAQRNSOGAHLQRRGVPGGPGSDPTSRGCVASGWNPAILRALDPYKPRRAGFG 199
QY 200 ESRSRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPISIL 259
Db 200 ESRSRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPISIL 259
QY 260 NPINIVVVKVLLLRDSDSGPKVTGNAALTLRNFCAQWKLNKVDKHPYWDTAILFTRQ 319
Db 260 NPINIVVVKVLLLRDSDSGPKVTGNAALTLRNFCAQWKLNKVDKHPYWDTAILFTRQ 319
QY 320 DLGCAATCDPLGADVTGCDPKRSCSVIEDDGLPSAETTAHELGHVFMHPDNVKNVCEE 379
Db 320 DLGCAATCDPLGADVTGCDPKRSCSVIEDDGLPSAETTAHELGHVFMHPDNVKNVCEE 379
QY 380 VFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDGSHGDCLLDQSPKISLPEDLP 439
Db 380 VFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDGSHGDCLLDQSPKISLPEDLP 439
QY 440 ASYTLQOCELAFGCVSKPCPYMOYCTKLWCTCAKAGQWVCOTRHPFHWADGTSCEGKLC 499
Db 440 ASYTLQOCELAFGCVSKPCPYMOYCTKLWCTCAKAGQWVCOTRHPFHWADGTSCEGKLC 499
QY 500 LKACVERHNLKRRYDVGSMKAWDPYPCGSRCTCGGVQLARRQCTNPTPANGKYCEGVR 559
Db 500 LKACVERHNLKRRYDVGSMKAWDPYPCGSRCTCGGVQLARRQCTNPTPANGKYCEGVR 559
QY 560 KYRSCNLEPCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPDRCKLI 619
Db 560 KYRSCNLEPCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPDRCKLI 619
QY 620 CRANGTGYFVVLAPKVVVDGTLQSPDSTSVCGQKCIKAGCDGNLGSKRFRKCGVCGGDN 679
Db 620 CRANGTGYFVVLAPKVVVDGTLQSPDSTSVCGQKCIKAGCDGNLGSKRFRKCGVCGGDN 679
QY 680 KSKKVTGLTKPMHGYNFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNG 739
Db 680 KSKKVTGLTKPMHGYNFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNG 739
QY 740 HFVYSAVERDLVVKGLLRYSGTGAVESLQASRPTEPLTVEVLSVKMTPPRVRYSFY 799
Db 740 HFVYSAVERDLVVKGLLRYSGTGAVESLQASRPTEPLTVEVLSVKMTPPRVRYSFY 799
QY 800 LPKEPREDKSSHPKDRGSLVNSVLSNOVEOPDPPARWAGSWGSCSASGSGL 859
Db 800 LPKEPREDKSSHPKDRGSLVNSVLSNOVEOPDPPARWAGSWGSCSASGSGL 859
QY 860 QKRAVDCRGSAQRTVPACDAARHPVETQACGEPCTWELSAWSPCSKSGRFGRRSLK 919
Db 860 QKRAVDCRGSAQRTVPACDAARHPVETQACGEPCTWELSAWSPCSKSGRFGRRSLK 919
QY 920 CVHGGELLARDQCNLHRKPQELDFCVLRPC 950
Db 920 CVHGGELLARDQCNLHRKPQELDFCVLRPC 950

RESULT 3
AAU74751
ID: AAU74751 standard; Protein; 952 AA.
XX AC AAU74751;
XX DT
XX XX 09-APR-2002 (first entry)
XX DE Human protease PRTS-11 protein sequence.
XX KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis.
XX XX Homo sapiens.

XX WO200198468-A2.
PN 27-DEC-2001.
XX 13-JUN-2001; 2001WO-US191178.
XX 16-JUN-2000; 2000US-2123336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-218212P.
PR 14-JUL-2000; 2000US-218946P.
XX (INCY-) INCYTE GENOMICS INC.
XX Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Walla NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX WPI; 2002-090437/12.
DR N-PSDB; ABK12894.
XX Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX Claim 1; Page 144-146; 177pp; English.
XX The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present protein sequence represents
CC the human protease PRTS-11 protein of the invention.
XX SQ Sequence 952 AA;
Query Match 70.8%; Score 673; DB 23; Length 952;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 793; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 20 EPREVVVPIRLDPPDINGRRYWRGPDGSDGLIFQITAFQEDFYHLTTPDAFLAPAF 79
Db 20 EPREVVVPIRLDPPDINGRRYWRGPDGSDGLIFQITAFQEDFYHLTTPDAFLAPAF 79
QY 80 STEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDFAAAYSLCGLRGAFYGAEVISPL 139
Db 80 STEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDFAAAYSLCGLRGAFYGAEVISPL 139
QY 140 PNASAPAAQRNSOGAHLQRRGVPGGPGSDPTSRGCVASGWNPAILRALDPYKPRRAGFG 199
Db 140 PNASAPAAQRNSOGAHLQRRGVPGGPGSDPTSRGCVASGWNPAILRALDPYKPRRAGFG 199
QY 200 ESRSRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPISIL 259
Db 200 ESRSRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPISIL 259
QY 260 NPINIVVVKVLLLRDSDSGPKVTGNAALTLRNFCAQWKLNKVDKHPYWDTAILFTRQ 319
Db 260 NPINIVVVKVLLLRDSDSGPKVTGNAALTLRNFCAQWKLNKVDKHPYWDTAILFTRQ 319
QY 320 DLGCAATCDPLGADVTGCDPKRSCSVIEDDGLPSAETTAHELGHVFMHPDNVKNVCEE 379
Db 320 DLGCAATCDPLGADVTGCDPKRSCSVIEDDGLPSAETTAHELGHVFMHPDNVKNVCEE 379

Db 320 DLGATTCDTLGMADVGTMCDFRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVVKCEE 379
 QY 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPEDLP 439
 Db 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPEDLP 439
 QY 440 ASYTLSSQCELAFLGVGSKPCPYMOYCTKLWCTGKAGQMVCTRHFPWADGTSCEGKLC 499
 Db 440 ASYTLSSQCELAFLGVGSKPCPYMOYCTKLWCTGKAGQMVCTRHFPWADGTSCEGKLC 499
 QY 500 LKGACVERHNLKHRVDGSKWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVR 559
 Db 500 LKGACVERHNLKHRVDGSKWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVR 559
 QY 560 VKYRSCNLEPCPSASGKSPREQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619
 Db 560 VKYRSCNLEPCPSASGKSPREQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619
 QY 620 CRANGTGYFYVLAPK-VVDGTLCSPDSTSVCGQKICAKGCDNGLSKKRPDKGVCVCGD 678
 Db 620 CRANGTGYFYVLAPKVVVDGTLCSPDSTSVCGQKICAKGCDNGLSKKRPDKGVCVCGD 679
 QY 679 NKCKKVTGLFTKPMHGYNFVAIPAAGASSIDIRQRYKGLIGDDNYLALNQSQKYLIN 738
 Db 680 NKCKKVTGLFTKPMHGYNFVAIPAAGASSIDIRQRYKGLIGDDNYLALNQSQKYLIN 739
 QY 739 GHFVSAVERDLVKGSLRYSGTGAVERSLQASRPILFVEVLSVGMKTPPRVRYSF 798
 Db 740 GHFVSAVERDLVKGSLRYSGTGAVERSLQASRPILFVEVLSVGMKTPPRVRYSF 799
 QY 799 YLPKEPREDKSSHP 812
 Db 800 YLPKEPREDKSSHP 813

RESULT 4
 AAU72899
 ID AAU72899 standard; Protein; 928 AA.

XX AAU72899;
 XX
 DT 26-FEB-2002 (first entry)
 XX Human metalloprotease partial protein sequence #11.
 DE
 XX Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
 KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
 KW hypertensive; neuroleptic; neuroprotective; anabolic;
 KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;
 KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;
 KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
 KW immune-related disease; cardiovascular disease; neuronal disease;
 KW migraine; sexual dysfunction; mood disorder; attention disorder;
 KW cognition disorder; hypertension; hypertention; psychotic disorder;
 KW dyskinesia; metabolic disorder; inflammatory disorder.

XX Homo sapiens.
 XX Wo200183782-A2.
 PN
 XX
 PD 08-NOV-2001.
 XX
 PF 04-MAY-2001; 2001WO-US14431.
 XX
 PR 04-MAY-2000; 2000US-201879P.
 XX
 PA (SUGEN-) SUGEN INC.
 XX
 PI Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 PI Payne V;
 XX
 DR WPI; 2002-041502/05.
 DR N-PSDB; AAS97182.

XX Novel protease polypeptide useful for screening for substances that may
 PT be used to treat, e.g., cancers, immune-related diseases,
 PT cardiovascular disease, migraine, pain, psychotic and inflammatory
 PT disorders -
 XX Claim 28; Figure 2G; 232pp; English.
 PS
 XX The invention relates to an isolated, enriched, or purified protease
 CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
 CC screen for substances (S) that may modulate its activity. Administering
 CC S (which modulates protease activity in vitro) may be used to treat a
 CC disease or disorder selected from cancers (e.g., of tissues, of blood or
 CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
 CC brain, ovarian, bladder or kidney), immune-related diseases and
 CC disorders, cardiovascular disease, brain or neuronal-associated diseases
 CC (e.g., central or peripheral nervous system diseases, migraine, pain,
 CC sexual dysfunction, mood disorders, attention disorders, cognition
 CC disorders, hypotension, hypertension, psychotic disorders, neurological
 CC disorders and dyskinesias), metabolic disorders and inflammatory
 CC disorders. (I) may also be useful as a diagnostic tool for a disease or
 CC disorder such as those above. AAU72876-AAU72910 represent human
 CC protease amino acid sequences of the invention.
 XX
 SQ Sequence 928 AA;

Query Match 66.8%; Score 635; DB 23; Length 928;
 Best Local Similarity 99.9%; Pred No. 0;
 Matches 755; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 20 EPEREVVPRLDPDINGRRYYWRGPDGSDGLIFQITAFQEDPYLHLPDAQFLAPAF 79
 Db 45 EPEREVVPRLDPDINGRRYYWRGPDGSDGLIFQITAFQEDPYLHLPDAQFLAPAF 104
 QY 80 STEHLGVPLQGLTGGSSDLRRCFYSDVNAEPDSPAAYSLGGLRGAFYGAEVVISPL 139
 Db 105 STEHLGVPLQGLTGGSSDLRRCFYSDVNAEPDSPAAYSLGGLRGAFYGAEVVISPL 164
 QY 140 PNASAPAAQRNSQGAHLQRRGVGPGSDPTSRGVSAGWNPAILRALDPYKPRRAGFG 199
 Db 165 PNASAPAAQRNSQGAHLQRRGVGPGSDPTSRGVSAGWNPAILRALDPYKPRRAGFG 224
 QY 200 ESRRRRSRGRKRFVSIPIRYVETLVVADBSMVKFGADLEHYLLTLATAARLVRHSIL 259
 Db 225 ESRRRRSRGRKRFVSIPIRYVETLVVADBSMVKFGADLEHYLLTLATAARLVRHSIL 284
 QY 260 NPINIVVVKVLLLRDRSDGPKVTGNAALTLRNFCAWQKLNKVSOKHPEYWDTAILFTRQ 319
 Db 285 NPINIVVVKVLLLRDRSDGPKVTGNAALTLRNFCAWQKLNKVSOKHPEYWDTAILFTRQ 344
 QY 320 DLGATTCDTLGMADVGTMCDFRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVVKCEE 379
 Db 345 DLGATTCDTLGMADVGTMCDFRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVVKCEE 404
 QY 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPEDLP 439
 Db 405 VFGKLRANHMSPPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPEDLP 464
 QY 440 ASYTLSSQCELAFLGVGSKPCPYMOYCTKLWCTGKAGQMVCTRHFPWADGTSCEGKLC 499
 Db 465 ASYTLSSQCELAFLGVGSKPCPYMOYCTKLWCTGKAGQMVCTRHFPWADGTSCEGKLC 524
 QY 500 LKGACVERHNLKHRVDGSKWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVR 559
 Db 525 LKGACVERHNLKHRVDGSKWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVR 584
 QY 560 VKYRSCNLEPCPSASGKSPREQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619
 Db 585 VKYRSCNLEPCPSASGKSPREQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 644
 QY 620 CRANGTGYFYVLAPK-VVDGTLCSPDSTSVCGQKICAKGCDNGLSKKRPDKGVCVCGD 678
 Db 645 CRANGTGYFYVLAPKVVVDGTLCSPDSTSVCGQKICAKGCDNGLSKKRPDKGVCVCGD 704

Qy	200	ESRSRRSGRAKRFVSIPIRYVETLLVVADESVMKPHGADLEHYLLTLATAARLYRHP	SIL	259
Db	200	ESRSRRSGRAKRFVSIPIRYVETLLVVADESVMKPHGADLEHYLLTLATAARLYRHP	SIL	259
Qy	260	NPINIVVVKVLLLRDSDGPKVTGNAALTNRNFCAMOKLNKYSKDKHPEYWDTAILFTRQ	319	
Db	260	NPINIVVVKVLLLRDSDGPKVTGNAALTNRNFCAMOKLNKYSKDKHPEYWDTAILFTRQ	319	
Qy	320	DLCGATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTAHEL	G	364
Db	320	DLCGATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTAHEL	G	364
RESULT 6				
	AAE22540			
ID	AAE22540 standard; Protein: 321 AA.			
XX	AAE22540;			
AC	XX			
XX	XX			
DT	26-JUL-2002	(first entry)		
XX	XX			
DE	Human protease #1.			
XX	XX			
KW	Human; novel human protein; NHP; protease; biological disorder; obesity;			
KW	high blood pressure; arthritis; connective tissue disorder; infertility;			
KW	gene therapy; enzyme.			
XX	XX			
OS	Homo sapiens.			
XX	XX			
PN	WC200226949-A2.			
XX	XX			
PD	04-APR-2002.			
XX	XX			
PF	27-SEP-2001; 2001WO-US30350.			
XX	XX			
PR	29-SEP-2000; 2000US-236689P.			
XX	XX			
PA	(LEXI-) LEXICON GENETICS INC.			
XX	XX			
PT	Friddle CJ, Hilbun E;			
XX	XX			
DR	WPI; 2002-372123/40.			
DR	N-PSDB; AAD35568.			
XX	XX			
PT	Novel nucleic acid encoding a human protease, useful as a hybridization			
PT	probe for screening libraries and assessing gene expression patterns -			
XX	XX			
PS	Claim 5; Page 34-35; 41pp; English.			
XX	XX			
CC	The present sequence is novel human protein (NHP), human protease.			
CC	NHPs share structural similarity with animal proteases particularly			
CC	zinc metalloproteases. Sequences of the invention are useful in			
CC	therapeutic, diagnostic and pharmacogenomic applications. NHP			
CC	polynucleotides are used as hybridization probes for screening			
CC	libraries and assessing gene expression patterns. They can also be			
CC	used for treating related biological disorders such as obesity, high			
CC	blood pressure, arthritis, connective tissue disorders and infertility.			
CC	They are also used in gene therapy.			
XX	XX			
SQ	Sequence	321 AA;		
Query Match				
Best Local Similarity 31.6%; Score 300; DB 23; Length 321;				
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	20	EPREVVPVRLDPDINGRYYWRGPDSDGGLIFQITAFQEDFYHLTPDQAF	LAPAF	79
Db	20	EPREVVPVRLDPDINGRYYWRGPDSDGGLIFQITAFQEDFYHLTPDQAF	LAPAF	79
Qy	80	STELHGVPLQGLTGGSSDLRRCFYSGDYNABPDSFAAVSLCGGLRGAFGYRGA	YVISPL	139
Db	80	STELHGVPLQGLTGGSSDLRRCFYSGDYNABPDSFAAVSLCGGLRGAFGYRGA	YVISPL	139

CC	migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC	and autoimmune diseases. They can be used to treat patients afflicted
CC	with an invasive tumour, a brain tumour or brain injury.
xx	
SQ	Sequence 381 AA:
	Query Match 18.6%; Score 177; DB 21; Length 381;
	Best Local Similarity 100.0%; Pred. No. 3.le-168;
	Matches 177; Conservative 0; Mismatches 0; Indels 0 Gaps 0
QY	255 HPSILNPINIVVVKVLLLRDRDSGPKVTGNAAALTLRNCAMQKLNKVDKHPEYWDTAI 314
Db	16 HPSILNPINIVVVKVLLLRDRDSGPKVTGNAAALTLRNCAMQKLNKVDKHPEYWDTAI 75
QY	315 LFTRQDLGGATCTDTLGMADVGTCMDPKRSCSVTEDDGLPSAFTTAEHLGHVFNNPHDNY 374
Db	76 LFTRQDLGGATCTDTLGMADVGTCMDPKRSCSVTEDDGLPSAFTTAEHLGHVFNNPHDNY 135
QY	375 KVCEEYFGKLRAHHMMSPTLIQIDRANPWSACSAAIITDFLDSGHGDCLLDQPSKPI 431
Db	136 KVCEEYFGKLRAHHMMSPTLIQIDRANPWSACSAAIITDFLDSGHGDCLLDQPSKPI 192

RESULT 8	
AAB21257	AAB21257 standard; Protein; 505 AA.
XX	
XX	
AC	AAB21257;
XX	
DT	23-FEB-2001 (first entry)
XX	
DE	Rat metalloproteinase ADAMTS-5.
XX	
KW	Rat; ADAMTS-5; metalloproteinase; ADAM;
KW	a disintegrin and metalloproteinase domain; thrombospondin domain;
KW	vaccine; neurotropic; neuroprotective; antiparkinsonian;
KW	cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW	Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
XX	autoimmune disease; brain tumour; brain injury.
XX	
OS	Rattus norvegicus.
XX	
XX	WO200053774-A2.
PN	
XX	
PD	14-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US06237.
XX	
PR	08-MAR-1999; 99US-0264585.
XX	
XX	(NEUR-) NEUROCRINE BIOSCIENCES INC.
PA	
XX	
PI	Keiner GS, Clark M, Maki RA;
XX	
XX	WPI; 2000-594326/56.
DR	N-PSDB; AAA95827.
DR	
XX	
XX	Polynucleotide encoding novel members of a disintegrin,
PT	metalloproteinase and thrombospondin domain protein family used to
PT	prevent and treat Alzheimer's disease, cancer and autoimmune diseases
XX	
XX	Claim 12; Fig 14; 129pp; English.
PS	
XX	

The present sequence is rat metalloproteinase ADAMTS-5. The ADAMTS family of proteins is closely related to the ADAM (A Disintegrin and Metalloproteinase Domain) family. Members of the ADAMTS family contain a thrombospondin domain in addition to the disintegrin and metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are useful for the manufacture of medicaments for treating conditions associated with neuroinflammation and/or neurodegeneration, such as Alzheimer's disease, Parkinson's disease and stroke. They are also useful for treating conditions associated with cell proliferation, cell migration, inflammation and/or angiogenesis, such as cancer arthritis

CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.

XX Sequence 505 AA;

Query Match 16.2%; Score 154; DB 21; Length 505;

Best Local Similarity 100.0%; Pred. No. 4.3e-145;

Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GPKVTGNAALTNRFCWOKKLNKVSDEKPEYWDTAIFTRDGLCGATTCDTLGMADVGT 337

Db 202 GPKVTGNAALTNRFCWOKKLNKVSDEKPEYWDTAIFTRDGLCGATTCDTLGMADVGT 261

QY 338 MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNKVKCEEVFGKLRANHMSPPTLIQI 397

Db 262 MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNKVKCEEVFGKLRANHMSPPTLIQI 321

QY 398 DRANPWSACSAAIITDFLDSGHGDCILLDQPSKPI 431

Db 322 DRANPWSACSAAIITDFLDSGHGDCILLDQPSKPI 355

RESULT 9

AAW78435

ID AAW78435 standard; Protein; 727 AA.

XX AC AAW78435;

XX 11-MAY-1999 (first entry)

XX Human ADAMTS-1 protein.

XX Metalloproteinase-disintegrin protein with a thrombospondin domain; ARDS;
KW ADAMTS-1; drug composition; foodstuffs; leukocyte; thrombocyte; hepatitis;
KW blood; erythrocyte; inflammatory disease; rheumatoid arthritis; asthma;
KW nephritis; Crohn's disease; acute respiratory disease syndrome.

XX Homo sapiens.

XX WO9855643-A1.

XX 10-DEC-1998.

XX 03-JUN-1998; 98WO-JP02449.

XX 03-JUN-1997; 97JP-0160422.

XX (KURE) KUREHA CHEM IND CO LTD.

XX Hakezaki M, Hirose K, Inouchi E, Ishida Y, Ishioka K;

PI Kuno K, Matsushima K;

XX WPI: 1999-070277/06.

DR N-PSDB; AAX17990.

XX Human metalloproteinase-disintegrin protein with thrombospondin

PT domain - useful as leukocyte and thrombocyte decreasing and

PT erythrocyte increasing agent

PS Claim 1; Page 51-52; 82pp; Japanese.

XX This sequence represents a novel human metalloproteinase-disintegrin
CC protein with a thrombospondin domain (ADAMTS-1). The protein may be used
CC in drug compositions and foodstuffs, as an agent for decreasing the
CC leukocyte and thrombocyte blood count and increasing the erythrocyte
CC blood count, e.g. for treatment of inflammatory diseases such as
CC rheumatoid arthritis, hepatitis, nephritis, Crohn's disease, asthma
CC and ARDS.

XX Sequence 727 AA;

Query Match

Best Local Similarity 100.0%; Score 17; DB 20; Length 727;

Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372

Db 156 AFTTAHELGHVFNMPHD 172

RESULT 10

AAW21265

ID AAW21265 standard; Protein; 896 AA.

XX AC AAW21265;

XX 23-FEB-2001 (first entry)

XX Mouse metalloproteinase ADAMTS-1.

XX Mouse; ADAMTS-1; metalloproteinase; ADAM;

KW a disintegrin and metalloproteinase domain; thrombospondin domain;

KW vaccine; nontropic; neuroprotective; antiparkinsonian;

KW cerebroprotective; cytostatic; antiarthritic; immunosuppressive;

KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;

KW autoimmune disease; brain tumour; brain injury.

XX Mus musculus.

XX WO200053774-A2.

XX 14-SEP-2000.

XX 08-MAR-2000; 2000WO-US06237.

XX 08-MAR-1999; 99US-0264585.

XX (NEUR-) NEUROCRINE BIOSCIENCES INC.

XX Kelner GS, Clark M, Maki RA;

XX WPI: 2000-594326/56.

XX Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Disclosure; Fig 17; 129pp; English.

XX The present sequence is mouse metalloproteinase ADAMTS-1. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.

XX Sequence 896 AA;

Query Match

Best Local Similarity 100.0%; Score 17; DB 21; Length 896;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372

Db 380 AFTTAHELGHVFNMPHD 396

RESULT 11

AAW49501

ID AAW49501 standard; Protein; 950 AA.

```
XX AC AAY49501;
XX DT 10-JAN-2000 (first entry)
XX DE Human METH1 protein.
XX KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
XX KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
XX KW angiogenesis inhibitor; abnormal wound healing; inflammation;
XX KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
XX KW diabetic retinopathy; macula degeneration; haemangioma; detection;
XX KW arterial-venous malformation; immune deficiency.
XX OS Homo sapiens.
XX PN WO9937660-A1.
XX PD 29-JUL-1999.
XX PF 22-JAN-1999; 99WO-0501313.
XX PR 23-JAN-1998; 98US-0072298.
XX PR 28-AUG-1998; 98US-0098539.
XX PA (IRUE/) IRUELA-ARISPE L.
XX PA (HAST/) HASTINGS G A.
XX PA (RUBE/) RUBEN S M.
XX PI IrueLa-Arispe L, Hastings GA, Ruben SM;
XX DR WPI; 1999-590684/50.
XX DR N-PSDB; AA232000.
XX PS New isolated metalloprotease thrombospondin polypeptides, useful for
XX PS treating hyperproliferative disorders, cancers or autoimmune disorders
XX PS Claim 10; Fig 1; 457pp; English.
XX CC AA232000 and AA232001 encode, and AAY49501 and AAY49502 represent, human
XX CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
XX CC respectively. METH1 and METH2 have been found to be potent inhibitors of
XX CC angiogenesis both in vitro and in vivo. They can be used for treating
XX CC cancer and other disorders related to angiogenesis including abnormal
XX CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
XX CC endometrial bleeding disorders, diabetic retinopathy, some forms of
XX CC macula degeneration, haemangiomas, and arterial-venous malformations.
XX CC They may be useful in treating deficiencies or disorders of the immune
XX CC system, by activating or inhibiting the proliferation, differentiation,
XX CC or mobilisation (chemotaxis) of immune cells. The etiology of these
XX CC immune deficiencies or disorders may be genetic, somatic, such as
XX CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
XX CC toxins), or infectious. They can also be used to treat inflammatory
XX CC conditions, both chronic and acute conditions. The products can also be
XX CC used for detection and diagnosis. AA232002 to AA232080, and AAY49503 to
XX CC AAY49511 represent sequences given in the exemplification of the present
XX CC invention.
XX SQ Sequence 950 AA;
Query Match 1.8%; Score 17; DB 20; Length 950;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 356 AFTTAHELGHVFNPHD 372
DB 379 AFTTAHELGHVFNPHD 395
RESULT 12
AAY53899
ID AAY53899 standard; Protein; 950 AA.
XX AC AAY53899;
XX DT 13-MAR-2000 (first entry)
XX DE Amino acid sequence of a murine ADAMTS-1 protein.
XX KW GON-1; metalloprotease; cell migration; modulator; ADAMTS-1;
XX KW metalloprotease domain; thrombospondin domain; abnormal cell migration;
XX KW organ shaping; sterility; cancer metastasis.
XX OS Mus sp.
XX PN WO9961656-A2.
XX PD 02-DEC-1999.
XX PF 28-MAY-1999; 99WO-US11918.
XX PR 29-MAY-1998; 98US-0087170.
XX PR 13-APR-1999; 99US-0129023.
XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.
XX PI Kimble JE, Bielloch RH;
XX DR WPI; 2000-072633/06.
XX PS Identifying modulators of proteins containing metalloprotease and
XX PS thrombospondin domains, potentially useful for controlling cell
XX PS migration and organ shaping -
XX PS Disclosure; Fig 1C; 60pp; English.
XX CC The present sequence represents a murine ADAMTS-1 protein. ADAMTS-1 is
XX CC a metalloprotease. The specification describes another related
XX CC metalloprotease, a C. elegans GON-1 protein, that lacks a transmembrane
XX CC domain and possesses a predicted metalloprotease domain between residues
XX CC 269-456. In C. elegans hermaphrodites, GON-1 is required for migration of
XX CC two distal tip cells to produce elongated tubes, whereas in males, GON-1
XX CC is required for migration of a single linker cell to produce a single
XX CC elongated tube. The protein is used in the method of the invention. The
XX CC specification describes a method for identifying a modulator of a
XX CC protein that contains a metalloprotease domain and a thrombospondin
XX CC domain. The method comprises treating a target organism, having a
XX CC developing gonadal cell that is responsive to the protein, with a test
XX CC compound, and determining any change in migration or shape of the cell
XX CC attributable to the test compound. The compounds identified are
XX CC potential therapeutic modulators of abnormal cell migration and organ
XX CC shaping, e.g. for rendering animals (specifically nematodes) sterile
XX CC and for inhibiting cancer metastases.
XX SQ Sequence 950 AA;
Query Match 1.8%; Score 17; DB 21; Length 950;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 356 AFTTAHELGHVFNPHD 372
DB 379 AFTTAHELGHVFNPHD 395
RESULT 13
AAB73549
ID AAB73549 standard; Protein; 950 AA.
XX AC AAB73549;
XX DT 07-AUG-2001 (first entry)
XX DE Human ADAM-type metalloprotease MDTs4, SEQ ID NO:4.
XX
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:20:53 ; Search time 24 Seconds
(without alignments)
1164.657 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 950
Sequence: 1 MLLGILTLFAGRTAGGFE.....DQCNLHRKPOELDFCVLRPC 950

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2.6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/1/iaa/PCUUS_COMB.pep.*
6: /cgn2.6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	1.8	608	4	US-09-130-491-13
2	17	1.8	967	4	US-09-130-491-2
3	14	1.5	551	4	US-09-130-491-16
4	13	1.4	837	4	US-09-122-126B-2
5	13	1.4	874	4	US-09-369-364A-15
6	13	1.4	1882	4	US-09-369-364A-13
7	12	1.3	481	4	US-09-130-491-8
8	12	1.3	905	4	US-09-369-364A-9
9	11	1.2	930	4	US-09-369-364A-2
10	11	1.2	930	4	US-09-122-126B-15
11	10	1.1	518	4	US-09-369-364A-22
12	10	1.1	1205	4	US-09-491-522-11
13	10	1.1	1211	4	US-09-491-522-5
14	10	1.1	1224	4	US-09-930-872-4
15	9	0.9	859	4	US-09-369-364A-5
16	9	0.9	1081	4	US-09-369-364A-17
17	8	0.8	16	4	US-09-369-364A-23
18	8	0.8	17	4	US-09-122-126B-13
19	8	0.8	26	4	US-09-122-126B-4
20	8	0.8	109	4	US-09-199-637A-25
21	8	0.8	148	3	US-08-946-329A-60
22	8	0.8	525	4	US-09-369-364A-21
23	8	0.8	788	2	US-08-918-914-4
24	7	0.7	10	1	US-07-801-812A-6
25	7	0.7	10	1	US-08-487-568-6
26	7	0.7	12	5	PCT-US93-03748-11
27	7	0.7	13	1	US-07-801-812A-18

28	7	0.7	13	1	US-08-487-568-18	Sequence 18, Appl
29	7	0.7	13	4	US-08-871-561-29	Sequence 29, Appl
30	7	0.7	13	4	US-09-321-932B-29	Sequence 29, Appl
31	7	0.7	23	1	US-07-646-531D-7	Sequence 7, Appl
32	7	0.7	23	2	US-08-488-273-7	Sequence 7, Appl
33	7	0.7	23	4	US-09-197-770B-13	Sequence 13, Appl
34	7	0.7	23	6	5426100-7	Patent No. 5426100
35	7	0.7	28	2	US-08-449-933-9	Sequence 9, Appl
36	7	0.7	28	4	US-07-966-049A-9	Sequence 9, Appl
37	7	0.7	56	2	US-08-557-309B-29	Sequence 29, Appl
38	7	0.7	56	3	US-08-834-306-29	Sequence 29, Appl
39	7	0.7	56	4	US-08-993-674A-29	Sequence 29, Appl
40	7	0.7	56	4	US-09-256-976-29	Sequence 29, Appl
41	7	0.7	106	2	US-08-557-309B-44	Sequence 44, Appl
42	7	0.7	106	3	US-08-834-306-44	Sequence 44, Appl
43	7	0.7	106	4	US-08-993-674A-44	Sequence 44, Appl
44	7	0.7	106	4	US-09-256-976-44	Sequence 44, Appl
45	7	0.7	111	4	US-08-899-330-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-130-491-13
; Sequence 13, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-130-491-13

Query Match 1.8%; Score 17; DB 4; Length 608;
Best Local Similarity 100.0%; Pred. No. 9.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
Db 91 AFTTAHELGHVFNMPHD 107
|||||

RESULT 2
US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2

Db 311 DTAILFTRQDLG 323

RESULT 7
US-09-130-491-8
; Sequence 8, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANG

FILE REFERENCE: 09404/041001
CURRENT APPLICATION NUMBER: US/09/130,491
CURRENT FILING DATE: 1998-08-07
EARLIER APPLICATION NUMBER: US 60/058,108
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: US 60/054,961
EARLIER FILING DATE: 1997-08-06
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 481
TYPE: PRT
ORGANISM: Rattus rattus
US-09-130-491-8

Query Match 1.3%; Score 12; DB 4; Length 481;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 603 WPKYSGVSPRD 614
|||||
Db 202 WPKYSGVSPRD 213

RESULT 8
US-09-369-364A-9
; Sequence 9, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-8
US-09-369-364A-9

Query Match 1.3%; Score 12; DB 4; Length 905;
Best Local Similarity 100.0%; Pred. No. 0.0081;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 603 WPKYSGVSPRD 614
|||||
Db 626 WPKYSGVSPRD 637

RESULT 9
US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

Query Match 1.2%; Score 11; DB 4; Length 930;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 CDTLGMADVGT 337
|||||
Db 376 CDTLGMADVGT 386

RESULT 10
US-09-122-126B-15
; Sequence 15, Application US/09122126B
; Patent No. 6451575
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-15

Query Match 1.2%; Score 11; DB 4; Length 930;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 CDTLGMADVGT 337
|||||
Db 376 CDTLGMADVGT 386

RESULT 11
US-09-369-364A-22
; Sequence 22, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-5
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (99)
; OTHER INFORMATION: Xaa = M
US-09-369-364A-22

Query Match 1.1%; Score 10; DB 4; Length 518;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 DKCGVCGDN 679
|||||
Db 307 DKCGVCGDN 316

RESULT 12
US-09-491-522-11
; Sequence 11, Application US/09491522
; Patent No. 6428998

```
; GENERAL INFORMATION:
; APPLICANT: Collige, Alain
; APPLICANT: Lapiere, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-491-522-11

Query Match 1.1%; Score 10; DB 4; Length 1205;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 DKCGVCGGDN 679
DB 704 DKCGVCGGDN 713
|||||

RESULT 13
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Collige, Alain
; APPLICANT: Lapiere, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
```

```
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-491-522-5

Query Match 1.1%; Score 10; DB 4; Length 1211;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 DKCGVCGGDN 679
DB 710 DKCGVCGGDN 719
|||||

RESULT 14
US-09-930-872-4
; Sequence 4, Application US/09930872
; Patent No. 6448388
; GENERAL INFORMATION:
; APPLICANT: Fiddie, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. 6448388 Human Proteases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0219-USA
; CURRENT APPLICATION NUMBER: US/09/930,872
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,852
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1224
; TYPE: PRT
; ORGANISM: homo sapiens
; US-09-930-872-4

Query Match 1.1%; Score 10; DB 4; Length 1224;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 527 PCSTRTCGGGV 536
DB 597 PCSTRTCGGGV 606
|||||

RESULT 15
US-09-369-364A-5
; Sequence 5, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
```

```

; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-6
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (450)
; OTHER INFORMATION: Xaa = L
US-09-369-364A-5

```

```

Query Match          0.9%; Score 9; DB 4; Length 859;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 528 CSRTCGGV 536
    |||||
Db 522 CSRTCGGV 530

```

Search completed: April 29, 2003, 17:24:27
Job time : 26 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:25 ; Search time 35.9135 Seconds
(without alignments)
2738.218 Million cell updates/sec

Title: US-10-009-332-1_copy_213_950

Perfect score: 4043

Sequence: 1 FVSPRYVETLVVADESMVK.....DQNLHRKPOELDFCVLRPC 738

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4043	100.0	950	22 AAG62299	Human metalloprote.
2	4043	100.0	950	23 AAE22541	Human protease #2.
3	3998	98.9	952	23 AAU74751	Human protease PRT
4	3743	5	92.6	23 AAU72899	Human metalloprote
5	2260	55.9	727	20 AAU78435	Human ADAMTS-1 pro
6	2254	55.8	950	20 AAU49501	Human METH1 protei
7	2254	55.8	950	22 AAU73549	Human ADAM-type me
8	2254	55.8	950	22 AAB50002	Human METH1. Homo
9	2254	55.8	967	19 AAW80285	Human integrin lig
10	2254	55.8	967	20 AAU04142	Human Tango-71 pro

11	2254	55.8	968	22 AAB50011	Protein; SEQ ID 12
12	2241	55.4	967	20 AAW78189	Human secreted pro
13	2220	54.9	896	21 AAY53899	Amino acid sequenc
14	2096.5	51.9	950	21 AAB21265	Mouse metalloprote
15	1966	48.6	381	21 AAB21261	Human metalloprote
16	1920.5	47.5	505	21 AAB21257	Rat metalloprote
17	1911.5	47.3	905	22 AAB72284	Murine ADAMTS-8 am
18	1900	47.0	890	20 AAY49502	Human METH2 protei
19	1900	47.0	890	22 AAB50003	Human METH2. Homo
20	1891	46.8	889	22 AAB74946	Human ADAM type me
21	1888	46.7	680	21 AAB21251	Human metalloprote
22	1779	44.0	1629	23 AAG30703	Human aggrecanase
23	1779	44.0	1629	23 AAG14448	Human ADAMTS-SI pr
24	1779	44.0	1916	23 AAE19173	Human protease, PR
25	1779	44.0	1935	23 AAG72896	Human metalloprote
26	1774	43.9	1602	23 AAG30702	Human aggrecanase
27	1761	43.6	837	20 AAW75425	Human aggrecan deg
28	1761	43.6	837	22 AAG78228	Human metalloprote
29	1761	43.6	840	21 AAB21256	Human metalloprote
30	1760	43.5	837	21 AAY99429	Human PRO1563 (UNQ
31	1760	43.5	837	22 AAU29199	Human PRO polypept
32	1760	43.5	837	22 AAB66178	Protein of the inv
33	1757.5	43.5	1073	21 AAB21264	Human metalloprote
34	1757.5	43.5	1882	22 AAB72286	Human ADAMTS-9 am
35	1757.5	43.5	1934	22 AAB72301	Human ADAMTS-9 alt
36	1745	43.2	930	20 AAW75426	Human aggrecan deg
37	1740	43.0	929	21 AAB41226	Human OREF ORF990
38	1730.5	42.8	625	23 AAM48394	Rat aggrecanase.
39	1730.5	42.8	874	22 AAB72287	Murine ADAMTS-9 am
40	1726	42.7	930	22 AAB72280	Murine ADAMTS-5 am
41	1696	41.9	1907	23 AAU77133	Human protease #12
42	1686	41.7	947	22 AAB86950	Human metalloprote
43	1685	41.7	958	21 AAB21255	Human metalloprote
44	1664	41.2	1505	23 AAB72897	Human metalloprote
45	1644.5	40.7	870	21 AAB21252	Rat metalloprote

ALIGNMENTS

RESULT 1

AAG62299

ID AAG62299 standard; protein; 950 AA.

XX AC

XX AAG62299;

XX AC

XX 23-AUG-2001 (first entry)

XX DT

XX Human metalloprotease MDR56 protein.

XX DE

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

XX KW

XX osteopathic; antiarthritic.

XX KW

XX Homo sapiens.

XX OS

XX WO200134785-A1.

XX PD

XX 17-MAY-2001.

XX PF

XX 10-NOV-2000; 2000WO-JP07917.

XX XX

XX 11-NOV-1999; 99JP-0321740.

XX PR

XX 16-MAY-2000; 2000JP-0144020.

XX PA

XX (YAMA) YAMANOUCHI PHARM CO LTD.

XX PA

XX (KAZU-) KAZUSA DNA RES INST.

XX XX

XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX PI

XX WPI; 2001-343602/36.

XX DR

XX N-PSDB; AAH41003.

XX XX

XX Metalloprotease with aggrecanase activity for treating joint diseases

XX PT

PT especially osteoarthritis
XX
PS Claim 1; Page 56-60; 85pp; Japanese.
XX
XX This invention relates to a metalloprotease with aggreganase activity.
CC The invention includes protein and DNA sequences of the metalloprotease,
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The present sequence represents
CC osteopathic and antiarthritic activity. The present sequence represents
CC the metalloprotease of the invention termed MDTIS6.
XX
XX
XX Sequence 950 AA;
SQ
Query Match 100.0%; Score 4043; DB 22; Length 950;
Best Local Similarity 100.0%; Pred. No. 4.2e-311;
Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVSIPTVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLL 60
DB 213 FVSIPTVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLL 272
QY 61 RDRSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAIFLTRQDLGCGATTCDTLGM 120
DB 273 RDRSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAIFLTRQDLGCGATTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEEVFGKLRANHMSP 180
DB 333 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEEVFGKLRANHMSP 392
QY 181 TLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPI SLPEDLPASVYTL SQOCELAF 240
DB 393 TLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPI SLPEDLPASVYTL SQOCELAF 452
QY 241 GVSGKPCPYMOYCTKLWCTGKAKGMVYCOTRHPFWADGTSCGEGKLC LKGCACVERHNLN 300
DB 453 GVSGKPCPYMOYCTKLWCTGKAKGMVYCOTRHPFWADGTSCGEGKLC LKGCACVERHNLN 512
QY 301 HRVDGSKWAKWDPYGCSTRTGGVQLARRQCTNPANGKYCEGVVYKRSCLNLEPCPS 360
DB 513 HRVDGSKWAKWDPYGCSTRTGGVQLARRQCTNPANGKYCEGVVYKRSCLNLEPCPS 572
QY 361 SASGKSFREEOCEAFNGYNHSTNRLTLAVAVVPKYSGVSPRDKKLCICRANGTGYFYLA 420
DB 573 SASGKSFREEOCEAFNGYNHSTNRLTLAVAVVPKYSGVSPRDKKLCICRANGTGYFYLA 632
QY 421 PKVVDGTLCSPDSTVCVQGCIRKAGCDNGLSKRPDKCGVCGDNKSKKVTGLFTKP 480
DB 633 PKVVDGTLCSPDSTVCVQGCIRKAGCDNGLSKRPDKCGVCGDNKSKKVTGLFTKP 692
QY 481 MHGYNFVVAIPAGASSIDIRQYKGLIGDDNYLALKNSQKYL LNHGFVYSAVERDLVY 540
DB 693 MHGYNFVVAIPAGASSIDIRQYKGLIGDDNYLALKNSQKYL LNHGFVYSAVERDLVY 752
QY 541 KGLRLYSTGTAVESLQASRPILFPLAVEVLSVCKMTPPRVYFSLPKPEPREDKSSHP 600
DB 753 KGLRLYSTGTAVESLQASRPILFPLAVEVLSVCKMTPPRVYFSLPKPEPREDKSSHP 812
QY 601 KDPGRGSLVHNSVLSLSNQVEQDPPARVWAGSWGPCSCSGSLQKRAVDCRGSAQ 660
DB 813 KDPGRGSLVHNSVLSLSNQVEQDPPARVWAGSWGPCSCSGSLQKRAVDCRGSAQ 872
QY 661 RTVPACDAAHRPVETOACGEPCTWELSAWSPCSCSGRGFORRLKCVGHGGRLLARDQ 720
DB 873 RTVPACDAAHRPVETOACGEPCTWELSAWSPCSCSGRGFORRLKCVGHGGRLLARDQ 932
QY 721 CNLHRKPQELDFCVLRPC 738
DB 933 CNLHRKPQELDFCVLRPC 950
RESULT 2

AAE22541
ID AAE22541 standard; Protein; 950 AA.
XX
AC AAE22541;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #2.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX
OS Homo sapiens.
XX
PN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Friddle CJ, Hilbun E;
XX
PI WPI; 2002-372123/40.
XX
DR N-PSDB; AAD35569.
XX
XX Novel nucleic acid encoding a human protease, useful as a hybridization
PT probe for screening libraries and assessing gene expression patterns.
XX
PS Claim 6; Page 36-38; 4lpp; English.
XX
CC The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridisation probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 950 AA;
Query Match 100.0%; Score 4043; DB 23; Length 950;
Best Local Similarity 100.0%; Pred. No. 4.2e-311;
Matches 733; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVSIPTVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLL 60
DB 213 FVSIPTVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLL 272
QY 61 RDRSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAIFLTRQDLGCGATTCDTLGM 120
DB 273 RDRSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAIFLTRQDLGCGATTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEEVFGKLRANHMSP 180
DB 333 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEEVFGKLRANHMSP 392
QY 181 TLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPI SLPEDLPASVYTL SQOCELAF 240
DB 393 TLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPI SLPEDLPASVYTL SQOCELAF 452
QY 241 GVSGKPCPYMOYCTKLWCTGKAKGMVYCOTRHPFWADGTSCGEGKLC LKGCACVERHNLN 300
DB 453 GVSGKPCPYMOYCTKLWCTGKAKGMVYCOTRHPFWADGTSCGEGKLC LKGCACVERHNLN 512
QY 301 HRVDGSKWAKWDPYGCSTRTGGVQLARRQCTNPANGKYCEGVVYKRSCLNLEPCPS 360
DB 513 HRVDGSKWAKWDPYGCSTRTGGVQLARRQCTNPANGKYCEGVVYKRSCLNLEPCPS 572
QY 361 SASGKSFREEOCEAFNGYNHSTNRLTLAVAVVPKYSGVSPRDKKLCICRANGTGYFYLA 420
DB 573 SASGKSFREEOCEAFNGYNHSTNRLTLAVAVVPKYSGVSPRDKKLCICRANGTGYFYLA 632
QY 421 PKVVDGTLCSPDSTVCVQGCIRKAGCDNGLSKRPDKCGVCGDNKSKKVTGLFTKP 480
DB 633 PKVVDGTLCSPDSTVCVQGCIRKAGCDNGLSKRPDKCGVCGDNKSKKVTGLFTKP 692
QY 481 MHGYNFVVAIPAGASSIDIRQYKGLIGDDNYLALKNSQKYL LNHGFVYSAVERDLVY 540
DB 693 MHGYNFVVAIPAGASSIDIRQYKGLIGDDNYLALKNSQKYL LNHGFVYSAVERDLVY 752
QY 541 KGLRLYSTGTAVESLQASRPILFPLAVEVLSVCKMTPPRVYFSLPKPEPREDKSSHP 600
DB 753 KGLRLYSTGTAVESLQASRPILFPLAVEVLSVCKMTPPRVYFSLPKPEPREDKSSHP 812
QY 601 KDPGRGSLVHNSVLSLSNQVEQDPPARVWAGSWGPCSCSGSLQKRAVDCRGSAQ 660
DB 813 KDPGRGSLVHNSVLSLSNQVEQDPPARVWAGSWGPCSCSGSLQKRAVDCRGSAQ 872
QY 661 RTVPACDAAHRPVETOACGEPCTWELSAWSPCSCSGRGFORRLKCVGHGGRLLARDQ 720
DB 873 RTVPACDAAHRPVETOACGEPCTWELSAWSPCSCSGRGFORRLKCVGHGGRLLARDQ 932
QY 721 CNLHRKPQELDFCVLRPC 738
DB 933 CNLHRKPQELDFCVLRPC 950
RESULT 2

Db	873	GQTVFACDAAHRPVETQAGCEPCTWELSAWSPCKSGRGFRQRRLSKCVGHGRLLAR	932
QY	719	DCNLHRKPOELDFCVLRPC	738
Db	933	DCNLHRKPOELDFCVLRPC	952
RESULT 4			
AAU72899			
ID	AAU72899	standard; Protein; 928 AA.	
AC	AC		
XX	AAU72899;		
DT	DT		
XX	26-FEB-2002	(first entry)	
DE	DE		
XX	Human metalloprotease partial protein sequence #11.		
KW	Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;		
KW	vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;		
KW	hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;		
KW	anorectic; antiinflammatory; aspartyl protease; cysteine protease;		
KW	metalloprotease; serine protease; cancer; haematopoietic; breast; colon;		
KW	lung; prostrate; cervical; brain; ovarian; bladder; kidney; pain;		
KW	immune-related disease; cardiovascular disease; neuronal disease;		
KW	migraine; sexual dysfunction; mood disorder; attention disorder;		
KW	cognition disorder; hypotension; hypertension; psychotic disorder;		
KW	dyskinesia; metabolic disorder; inflammatory disorder.		
OS	Homo sapiens.		
XX	XX		
FN	W0200183782-A2.		
XX	XX		
PD	08-NOV-2001.		
XX	XX		
PF	04-MAY-2001; 2001WO-US14431.		
XX	XX		
PR	04-MAY-2000; 2000US-201879P.		
XX	XX		
PA	(SUGE-) SUGEN INC.		
XX	XX		
PI	Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;		
PI	Payne V;		
DR	WPI; 2002-041502/05.		
DR	N-PSDB; RAAS9718B.		
XX	XX		
PT	Novel protease polypeptide useful for screening for substances that may		
PT	be used to treat, e.g., cancers, immune-related diseases,		
PT	cardiovascular disease, migraine, pain, psychotic and inflammatory		
PT	disorders -		
XX	XX		
PS	Claim 28; Figure 2G; 232pp; English.		
XX	XX		
CC	The invention relates to an isolated, enriched, or purified protease		
CC	polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to		
CC	screen for substances (S) that may modulate its activity. Administering		
CC	S (which modulates protease activity in vitro) may be used to treat a		
CC	disease or disorder selected from cancers (e.g., of tissues, of blood or		
CC	haematopoietic origin, of the breast, colon, lung, prostate, cervical,		
CC	brain, ovarian, bladder or kidney), immune-related diseases and		
CC	disorders, cardiovascular disease, brain or neuronal-associated diseases		
CC	(e.g., central or peripheral nervous system diseases, migraine, pain,		
CC	sexual dysfunction, mood disorders, attention disorders, cognition		
CC	disorders, hypotension, hypertension, psychotic disorders, neurological		
CC	disorders and dyskinesias), metabolic disorders and inflammatory		
CC	disorders. (I) may also be useful as a diagnostic tool for a disease or		
CC	disorder such as those above. AAU72876-AAU72910 represent human		
CC	protease amino acid sequences of the invention.		
XX	XX		
SQ	Sequence	928 AA;	
	Query Match	92.6%;	Score 3743.5; DB 23; Length 928;
	Best Local Similarity	93.4%;	Pred. No. 2.1e-287;

Db 714 VTSAPGCHDILITPTGATNIEVKQNRGRNNGSFLAIAKADGYTLNGDYLTLSTLEQ 773
 QY 537 DLVVKGSILRYSGTGTAVESLQASRPILPTVEVLVSKMTPPRVRYSFYLPKPREDK 596
 Db 774 DIMKGVVLYSGSSAALERIRSFPLKEPLTQVLTGVNLRPKIKYTFVKKK---K 829
 QY 597 SSHPKDPRGSPVLSLNSVLSNOVEQDDPPARWAGSWGPGCSASCSGLQKRAVDCRG 656
 Db 830 ES-----FNAIPTFS-----AWIEWEGECSKCELGWQRLVECRD 866
 QY 657 SAGQRTVPACDAH--RPVETOACGE-PCPTWELSAWSPCSKSGRFRRLSKVCVGHG 713
 Db 867 INGO---PASECAKEVKPASTRPCADHPCPQWOLGEWSSCSKTGGRYKRSILKCLSHDG 923
 QY 714 RLRLAQDQNLHRKPQE-LDFCVLRPC 738
 Db 924 GVLSHESCDPLKPKHFIDECTWAE 949

RESULT 8

AAB50002
 ID AAB50002 standard; Protein; 950 AA.
 AC AAB50002;
 XX
 DT 19-MAR-2001 (first entry)
 DE Human METH1.
 XX
 KW Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Webber syndrome;
 KW plaque neovascularisation; telangiectasia; haemophilic joint;
 KW angiofibroma; fibromuscular dysplasia; wound granulation;
 KW Crohn's disease; atherosclerosis; birth control.
 XX
 OS Homo sapiens.
 XX
 PN WO200071577-A1.
 PD 30-NOV-2000.
 XX
 PF 25-MAY-2000; 2000WO-US14462.
 XX
 PR 25-MAY-1999; 99US-0318208.
 PR 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONA/) JONAK Z L.
 PA (TRUL/) TRULLI S H.
 PA (FORN/) FORNWALD J A.
 PA (TERR/) TERRETT J A.
 XX
 PI IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Fornwald JA, Terrett JA;
 XX
 DR WPI; 2001-025136/03.
 DR N-PSDB; AAC90057.
 XX
 PT METH1 and METH2 polynucleotides and encoded polypeptides, used to

PT
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 XX rheumatoid arthritis and psoriasis -
 PS Claim 15; Fig 1; 768pp; English.
 XX
 CC The present sequence is human METH1 (ME for metalloprotease and TH for
 CC thrombospondin). METH1 can be used for inhibiting angiogenesis in an
 CC individual, and for treating cancer, benign tumours, an ocular angiogenic
 CC disease, rheumatoid arthritis, psoriasis, delayed wound healing, plaque
 CC endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion
 CC fractures, scleroderma, trachoma, vascular adhesions, myocardial
 CC angiogenesis, coronary collaterals, cerebral collaterals, arteriovenous
 CC malformations, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma, plaque
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH1 can also be used in birth control. METH1 can also
 CC be used in diagnostic methods for the prognosis of cancer.
 XX

SQ Sequence 950 AA;

Query Match 55.8%; Score 2254; DB 22; Length 950;
 Best Local Similarity 54.0%; Pred. No. 1.8e-169;
 Matches 403; Conservative 131; Mismatches 172; Indels 40; Gaps 11;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATLAARLYRHPHSILNPINIVVVKVLL 60
 Db 236 FVSSHRVYETMLVADQSMAEFHGSLKHYLLTLFSAARLYKHPHSIRNSVSLVVKVILI 295
 QY 61 RDRSGPKVTGNAALTLRNFCAWOKLKNVSKDHPXYDWTALFTRODLGATCTDLGM 120
 Db 296 HDEQKGPEVTSNAALTLRNFCNMQKQNPSPSDRAEHYDTALTFTRODLGSCQCTDLGM 355
 QY 121 ADVGTCMDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDNKVKCEVEFKLRANHMSP 180
 Db 356 ADVGTCMDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDADKACASLNGVNDSHMMAS 415
 QY 181 TLIQIDRANPWSACSAIITDLDGSHGDCLLDQSPISIPEDLPKASVTLSSOCELAF 240
 Db 416 MSLNLDHSQWSPCSAYNITSFLDNGHGECLMDKRPNIQIPGLPSTSDANHCQCOFTF 475
 QY 241 GVGSKPCP-YMQYCTKLWCTGKAKQMVCTRHFPFADGTSCGEGKLCGLKACVYERHNLN 299
 Db 476 GEDSKHCPDAASTCTLWCTGSGVLVQCTKHPPWADGTSCGEGKWCINGKCVNKTID-R 534
 QY 300 KH---RVDGSAKWDPYPCSRCTCGGVQVQLARQCTNPTPANGKCYCEGVYKRSCLNE 356
 Db 535 KHFTPPHGSWGMGMDCSRCTCGGVQVQTMRECDNPVKNKGKCYCEGKRVYRSCNLE 594
 QY 357 PCPSASGKSFREBQCEAFNGYNHSTNRLTLAVAWVPKYSVSPDRCKKLCICRANGTYF 416
 Db 595 DCPDN-NGKTFREBQCEAHNEFSKASFGSGPAVEMIPKYGAVSPKORCKLICQAKGIGYF 653
 QY 417 YVLAPKVVDTGLCSPDSTSVCGVKCIKACDGNLGSKRFPDCKGCVGGDNKSKCKVTGL 476
 Db 654 FVLPKVVDTGTPCSPDSTSVCGVQCVKAGCDRIIDSKKFKDCKGCVGGNGSTCKRISGS 713
 QY 477 FTKPMHGNFVVAIPAGASSIDIRQRYGKGLIGDNNYALKNSOGKYLNGHGVVSAVER 536
 Db 714 VTSAPGCHDILITPTGATNIEVKQNRGRNNGSFLAIAKADGYTLNGDYLTLSTLEQ 773
 QY 537 DLVVKGSILRYSGTGTAVESLQASRPILPTVEVLVSKMTPPRVRYSFYLPKPREDK 596
 Db 774 DIMKGVVLYSGSSAALERIRSFPLKEPLTQVLTGVNLRPKIKYTFVKKK---K 829
 QY 597 SSHPKDPRGSPVLSLNSVLSNOVEQDDPPARWAGSWGPGCSASCSGLQKRAVDCRG 656
 Db 830 ES-----FNAIPTFS-----AWIEWEGECSKCELGWQRLVECRD 866
 QY 657 SAGQRTVPACDAH--RPVETOACGE-PCPTWELSAWSPCSKSGRFRRLSKVCVGHG 713
 Db 867 INGO---PASECAKEVKPASTRPCADHPCPQWOLGEWSSCSKTGGRYKRSILKCLSHDG 923
 QY 714 RLRLAQDQNLHRKPQE-LDFCVLRPC 738

XX PS Claim 8; Fig 1; 84pp; English.

XX CC The present sequence represents human Tango-71. Tango polypeptides are

CC useful for identifying compounds which bind the polypeptide via direct

CC binding, competition binding assays or Tango-71, -73, -74, 76 or -83-

CC mediated signal transduction. Tango polypeptides are also useful for

CC identifying modulating compounds by determining effect on Tango activity.

CC Tango polypeptides and nucleic acids are useful for diagnosing diseases

CC related to aberrant expression of Tango, and Tango polypeptides are

CC useful for raising antibodies which can be used in diagnostic assays for

CC detection of Tango, and also for generating anti-idiotypic antibodies for

CC prevention and protection.

XX SQ Sequence 967 AA;

Query Match 55.8%; Score 2254; DB 20; Length 967;

Best Local Similarity 54.0%; Pred. No. 1.8e-169;

Matches 403; Conservative 131; Mismatches 172; Indels 40; Gaps 11;

QY 1 FVSIPRYVETLVVADESMVKFHGADLEHYLLTLLATAARLYRHPISILPINIVVVKVLL 60

DB 253 FVSSHRYVETMLVADQSAEFGHSGKHYLLTFLSVAARLYKHPISIRNSVSLVVKILVI 312

QY 61 RDRDSGPKVTGNAALTIRNFCWQKLNKYSKHPHYDWTAILFTRODLGATCTDITGM 120

DB 313 HDEQKQPEVTSNAALTIRNFCWQKLNKYSKHPHYDWTAILFTRODLGATCTDITGM 372

QY 121 ADVGTMCDPKRCSCVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEYFVKLRANHMSP 180

DB 373 ADVGTVCDFSRCSVIEDDGLQAFTTAHELGHVFNMPHDNAQKASLNGVNDQSHMMAS 432

QY 181 TLIQIDRANPWSACSAALITFDLSGHGDCLLDOPSKPISLPEDLPKASYTSLQOCELAF 240

DB 433 MSLNLDHQPWSPSCSAYMITSFELNGHGECLMDKPNQIPLPGDLPGTSYDANRQCOFTF 492

QY 241 GVGRKPCP-YMOYCTKLACTKAKQMYCQTRHPFADGTSCGSEKCLKAGACVERNNLN 299

DB 493 GEDSKHCPDAASTCTLMTSGSLVLCOTKHFPAWDTSGEGKWKINGKCNKTD-R 551

QY 300 KH--RVDSWAKWDPYPCSTCGGCVOLARROCTNTPANGKCYEGVVRKYRSCNLE 356

DB 552 KHFDPFHGSGWGPWGCSTCGGCVQYTWRECDNVPKNGKCYEGKRVYRSCNLE 611

QY 357 PCPSASGSKFEEQCEAFNGYHNHNTLTLAVAWPKYSGVSPDRCKLICRANGTGYF 416

DB 612 DCPDN-NGKTFREEQCEAHNEFSAFGSPAVEWPKYAGVSPKDRCKLICQAKGIFY 670

QY 417 VYLAQVVDGTLCPDSTSVCGQKICAKGCDNLGSKRFDKCGVCGDNKSKCKVTGL 476

DB 671 FVLQPKVVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKFDKCGVCGGNGSTCKKISGS 730

QY 477 FTKPMHGNFVVAIPAGASSIDIRQYKGLIGDNDVIALKNSOGKYLLNGHFVVSVAVER 536

DB 731 VTSAPKGYHDIITPTGATNIEVKORNGSRNNGSFLAIAAGDTYLLNGDYTLISLEQ 790

QY 537 DLVVKGSLRLSGTGAVESLOARPILEPLTVLSVGVKMTPPRVRYSFYLPKREPREDK 596

DB 791 DIMYKGVVLYRSGSAAALERTSRFSLKEPLTIQVLTWGNALRPKIKYTFVKKK---K 846

QY 597 SSHPKDPRGSLVNLVSLNSQVQEDDRPPARFWAGSWGPCASGSLGKRAVDCRG 656

DB 847 ES-----FNAIPFSS-----AWVIEWGECSKSGELGWQRRLVECRD 883

QY 657 SAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGFORSLKCVGHGG 713

DB 884 INGO---PASECAKEVKPASTRPCADHPCPOWQJGEWSSCKTKCGKFKRSLKCLSHDG 940

QY 714 RLLAQDQNLHRKPOE-LDFCVLRPC 738

DB 941 GVLSHESCDPLKPKHFDICTMAEC 966

RESULT 11

AAB50011

ID AAB50011 standard; Protein; 968 AA.

XX AC AAB50011;

XX DT 19-MAR-2001 (first entry)

XX DE Protein; SEQ ID 125.

XX KW Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;

KW cancer therapy; benign tumour; ocular angiogenic disease;

KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;

KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;

KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;

KW coronary collateral; cerebral collateral; arteriovenous malformation;

KW ischaemic limb angiogenesis; Osler-Webber syndrome;

KW plaque neovascularisation; telangiectasia; haemophilic joint;

KW angiofibroma; fibromuscular dysplasia; wound granulation;

KW Crohn's disease; atherosclerosis; birth control.

XX OS Homo sapiens.

XX PN WO200071577-A1.

XX PD 30-NOV-2000.

XX PF 25-MAY-2000; 2000WO-US14462.

XX PR 25-MAY-1999; 99US-0318208.

PR 20-JUL-1999; 99US-0144882.

PR 10-AUG-1999; 99US-0147823.

PR 13-AUG-1999; 99US-0373658.

PR 22-DEC-1999; 99US-0171503.

PR 22-FEB-2000; 2000US-0183792.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (SMK) SMITHKLINE BEECHAM CORP.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA (IRUE/) IRUELA-ARISPE L.

PA (HAST/) HASTINGS G A.

PA (RUBE/) RUBEN S M.

PA (JONA/) JONAK Z L.

PA (TRUL/) TRULLI S H.

PA (FORN/) FORNWALD J A.

PA (TER/) TERRETT J A.

PA IrueLa-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;

PI Fornwald JA, Terrett JA;

XX WPI; 2001-025136/03.

XX METH1 and METH2 polynucleotides and encoded polypeptides, used to

inhibit angiogenesis in the treatment of disorders such as cancer,

XX rheumatoid arthritis and psoriasis -

PS Claim 15; Pages 759-763; 768pp; English.

XX The present invention relates to human METH1 and METH2 (ME for

metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).

CC METH can be used for inhibiting angiogenesis in an individual, and for

CC treating cancer, benign tumours, an ocular angiogenic disease,

CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,

CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,

CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,

CC coronary collaterals, cerebral collaterals, arteriovenous malformations,

CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque

CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,

CC fibromuscular dysplasia, wound granulation, Crohn's disease or

CC atherosclerosis. METH can also be used in birth control. METH can also

CC be used in diagnostic methods for the prognosis of cancer. The present

sequence is a protein isolated in the present invention.

XX

CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. AAX04302) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 86 novel genes and their fragments (nucleic
CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 86
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX04311 for described uses).

XX SQ Sequence 967 AA;
Query Match 55.4%; Score 2241; DB 20; Length 967;
Best Local Similarity 53.8%; Pred. No. 2e-168;
Matches 401; Conservative 131; Mismatches 174; Indels 40; Gaps 11;
QY 1 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPIWVVKVLL 60
Db 253 FVSSHYVETMLVADQSAEFGSLGKHYLLTLFSAARLXKHPXIRNSVSLVVKVLLVI 312
QY 61 RDRDSGPKVTGNAALTFRNFCQAWKKLVSKHPEYWDTAILFTRODLGGATCDTLGM 120
Db 313 HDQKQPEVTSNAALTFRNFCQWQHNPSPDRDAEHYDTAILFTRODLGSGQTCDTLGM 372
QY 121 ADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNYKVEEFGKLRANHMSP 180
Db 373 ADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNYKVEEFGKLRANHMSP 180
QY 181 TLTIQIDRANPWSACSAIITFDLDSGHGDCILDQPSKIPSLPEDLPAGSYTLSQOCELAF 240
Db 433 MLNLDHSPQWSPCSAYMITSFLDNGHGBCLMDKQNPQLPGLDLPSTYDANROCOQTF 492
QY 241 GVSKECP- YMYCTKWLCTKAKQOMVCOTRHFWDGTSCEGKCLKLGACVYERHNLN 299
Db 493 GEDSKPCDAASTCSPLWCTGTSGVLVCOTRHFWDGTSCEGKCLKLGACVYERHNLN 299
QY 300 KH---RVDGSKAWDPYPCSTCGGVLARRQCTNPTTPANGKYCGVRYKYBSCNLE 356
Db 552 KHEPTFFHSGWGWPGWDCSTCGGVOYTWREDNPVFKNGKYCEKRYKSCNLE 611
QY 357 PCSSASGSKFRECEAFNGYNHSTNRLTLAVAWPKYSGVSPDKCLICRANGTYGF 416
Db 612 DCPDN-NGKTFRECEAHNEFSKASFGSPAVEMPKYAGVSPDKCLICRANGTYGF 670
QY 417 YVLAPKVDGTLCSPDSTSVCGQKICAGCDNGLSKRFPDKCGVCGGDNKSKKVTGL 476
Db 671 FVLQPKVDGTLCSPDSTSVCGQKICAGCDRIIDSKRFPDKCGVCGGDNKSKKISGS 730
QY 477 FTKPMHGYNFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQKYLINGHFVYSAYER 536
Db 731 VTSAPGYHDIITPGATNIEVKQNRQSRNGSFLAIAKAADGTYLNGDYTLSTLEQ 790
QY 537 DLVVGSLRYSGTAVESLQASRILEPLFVLSVSKMTPPRVYSFYLPKPREDK 596
Db 791 DIMYKGVLYRSGSSAALRIERSPLKEPLTIQVLTGVALRPIKITYFVKK- ---K 846
QY 597 SSHPKDPRGSPVLSVLSNQVEQDDRPAPRVAGSWGPCASCGSLQKRAVDCRG 656
Db 847 ES-----FNAIFTS-----AWIEWGCSKSCELGWQRRVLECRD 883
QY 657 SAGQRTVPACDAH---RPVETOACGE-PCPTWELSAWSPCSKSGRGFORSLKCVGHGG 713
Db 884 INQ---PASECAKEVKPASTRPCADHPCPQWQLGEWSSCSKTCCKGYKRSKLCLSHDG 940
QY 714 LLARQOCLNHRKPBQ-B-LDFCVLRPC 738
Db 941 GVLSHESCDPLKKRPHDFICTMAEC 966

RESULT 13

AAY53899
ID AAY53899 standard; Protein; 950 AA.
AC AAY53899;
XX AAY53899;
DT 13-MAR-2000 (first entry)
XX Amino acid sequence of a murine ADAMTS-1 protein.
DE GON-1; metalloproteinase; cell migration; modulator; ADAMTS-1;
KW metalloproteinase domain; thrombospondin domain; abnormal cell migration;
KW organ shaping; sterility; cancer metastasis.
XX Mus sp.
XX WO9961656-A2.
XX 02-DEC-1999.
XX 28-MAY-1999; 99WO-US11918.
XX 29-MAY-1998; 98US-0087170.
XX 13-APR-1999; 99US-0129023.
XX (WISC) WISCONSIN ALUMNI RES FOUND.
XX Kimble JE, Brelloch RH;
XX WPI; 2000-072633/06.
XX Identifying modulators of proteins containing metalloprotease and
XX thrombospondin domains, potentially useful for controlling cell
XX migration and organ shaping -
XX Disclosure; Fig 1C; 60pp; English.
XX The present sequence represents a murine ADAMTS-1 protein. ADAMTS-1 is
XX a metalloproteinase. The specification describes another related
XX metalloproteinase, a C. elegans GON-1 protein, that lacks a transmembrane
XX domain and possesses a predicted metalloprotease domain between residues
XX 269-436. In C. elegans hermaphrodites, GON-1 is required for migration of
XX two distal tip cells to produce elongated tubes, whereas in males, GON-1
XX is required for migration of a single linker cell to produce a single
XX elongated tube. The protein is used in the method of the invention. The
XX specification describes a method for identifying a modulator of a
XX protein that contains a metalloproteinase domain and a thrombospondin
XX domain. The method comprises treating a target organism, having a
XX developing gonadal cell that is responsive to the protein, with a test
XX compound, and determining any change in migration or shape of the cell
XX attributable to the test compound. The compounds identified are
XX potential therapeutic modulators of abnormal cell migration and organ
XX shaping, e.g. for rendering animals (specifically nematodes) sterile
XX and for inhibiting cancer metastases.

SQ Sequence 950 AA;

Query Match 54.9%; Score 2220; DB 21; Length 950;
Best Local Similarity 53.3%; Pred. No. 8e-167;
Matches 398; Conservative 130; Mismatches 177; Indels 42; Gaps 12;
QY 1 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPIWVVKVLL 60
Db 237 FVSSHYVETMLVADQSAEFGSLGKHYLLTLFSAARLYRHPSILNPIWVVKVLLVI 296
QY 61 RDRDSGPKVTGNAALTFRNFCQAWKKLVSKHPEYWDTAILFTRODLGGATCDTLGM 120
Db 297 YEOKQPEVTSNAALTFRNFCQWQHNPSPDRDAEHYDTAILFTRODLGSGQTCDTLGM 356
QY 121 ADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNYKVEEFGKLRANHMSP 180
Db 357 ADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNYKVEEFGKLRANHMSP 180
QY 181 TLTIQIDRANPWSACSAIITFDLDSGHGDCILDQPSKIPSLPEDLPAGSYTLSQOCELAF 240

Db 416 MLSSLDHSPWSPCSAYMVTSLFDNGHGCLMDKPNPIKLPDLFTGLTDANROCCFTF 475
QY 241 GVSGKPCP-YMOYCTKWLCTGKAGQWVQTRHFPWADGTSCGEGKCLKLGACVERHNLN 299
Db 476 GEESKPCPDAASTCTTLWCTGTSGGLVLCCTHFPWADGTSCGEGKWCVCVKNKTDN- 534
QY 300 KH---RVDSWAKWDYPGCSRTCGGVQVQLARRQCTNPTPANGKYCEGVRYKRSNLE 356
Db 535 KHFPATVHGSWGPWGPWGDSCRTCGGVQVYTMRECDNPVPKNGKYCEGKRYRSCNIE 594
QY 357 PCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAVPKYSGVSPDRCKLCIRANGTGYF 416
Db 595 DCPDN-NGKTFREOCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKLTCEAKGIGYF 653
QY 417 YVLAPKVVDDGTLCSPDSTVCGVQKCIKAGCDNLSKKRFDKCGVCGDNDKCKVVTGL 476
Db 654 FVLQPKVVDGTLCSPDSTVCGVQKCIKAGCDRIIDSKKKFDKCGVCGDNGSTCKKMSGI 713
QY 477 FTKPMHGYNFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNGHVFVSAYER 536
Db 714 VTSTRPGYHDIPTIPAGATNIEVKHNRQGRNNGSFLAIRAADGTYYILNGFTLSTLEQ 773
QY 537 DLVVKSLRYSGTGTAVESLQASRPILFTVEVLSVCKMPPRVRYSFYLPKPREDEK 596
Db 774 DLTYKGTVLRYSGSSAALERSFSLPEPLTIQVLMVGHALRPKIKFTYPMKKTES-- 831
QY 597 SSHPKDPRGSPVLSLHNSVLSNQVEQDDPPARWAGSWGPCSSASCGLQKRAVDRCG 656
Db 832 -----FNAIPTFS-----EWVIEWEGCSKTCGSGWQRRVYQCRD 867
QY 657 SAGQRTVPACDAH--RPVETOACGE-PCPTW 685
Db 867 INGH---PASECAKEVKPASTRPCADLPCPHW 896
QY 713 GRLLARDQCNLHRRKQPS-LDFCVLRPC 738
Db 924 GGVLSNESCDPLKKPKHYIDFCHLTQC 950
RESULT 14
AAB21265
ID AAB21265 standard; Protein; 896 AA.
XX AAB21265;
AC AAB21265;
DT 23-FEB-2001 (first entry)
XX Mouse metalloproteinase ADAMTS-1.
KW Mouse; ADAMTS-1; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; neurotropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytoskeletal; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX Mus musculus.
OS W0200053774-A2.
PN 14-SEP-2000.
XX 08-MAR-2000; 2000WO-US06237.
XX 08-MAR-1999; 99US-0264585.
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX Keiner GS, Clark M, Maki RA;
XX WPI; 2000-594326/56.

PT Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PS prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
PS Disclosure; Fig 17; 129pp; English.
XX The present sequence is mouse metalloproteinase ADAMTS-1. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.
XX Sequence 896 AA;
SQ Query Match 51.9%; Score 2096.5; DB 21; Length 896;
Best Local Similarity 54.2%; Pred. No. 4.8e-157;
Matches 375; Conservative 115; Mismatches 163; Indels 39; Gaps 9;
QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLTLLATARLARHPHSILNPINIVVVKVLL 60
Db 237 FVSSPRYVETMLVADQSMADFHGSLKHYLLTFSVAARFKHFSIRNSISLVVVKILVI 296
QY 61 RDRDSGPKVTGNAALTLNFCANQKKNVSDKHPEYWDATAILFTRODLGATTCDDTLGM 120
Db 297 YEEQKPEVTSNAALTLNFCNWKQHNPSDROPEHYDTAILFTRODLGSHCTDLGM 356
QY 121 ADVGTMCDPKRSCSVIEDGLPSAETTAHELGHVFNPMHNVKVEEFGKLRANHMSP 180
Db 357 ADVGTVCDPSRSCSVIEDGLQAAFTTAHELGHVFNMPHDDAKHACASLNGVTGSHLMA 416
QY 181 TLIQIDRANPWSACSAIITDLDSDHGDCLLDOPSKPISLPEDLPAGASYTLSSQCELA 240
Db 417 MLSSLDHSPWSPCSAYMVTSLFDNGHGCLMDKPNPIKLPDLFTGLTDANROCCFTF 476
QY 241 GVSGKPCP-YMOYCTKWLCTGKAGQWVQTRHFPWADGTSCGEGKCLKLGACVERHNLN 299
Db 477 GEESKPCPDAASTCTTLWCTGTSGGLVLCCTHFPWADGTSCGEGKWCVCVKNKTDN- 535
QY 300 KH---RVDSWAKWDYPGCSRTCGGVQVQLARRQCTNPTPANGKYCEGVRYKRSNLE 356
Db 535 KHFPATVHGSWGPWGPWGDSCRTCGGVQVYTMRECDNPVPKNGKYCEGKRYRSCNIE 595
QY 357 PCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAVPKYSGVSPDRCKLCIRANGTGYF 416
Db 595 DCPDN-NGKTFREOCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKLTCEAKGIGYF 654
QY 417 YVLAPKVVDDGTLCSPDSTVCGVQKCIKAGCDNLSKKRFDKCGVCGDNDKCKVVTGL 476
Db 654 FVLQPKVVDGTLCSPDSTVCGVQKCIKAGCDRIIDSKKKFDKCGVCGDNGSTCKKMSGI 714
QY 477 FTKPMHGYNFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNGHVFVSAYER 536
Db 714 VTSTRPGYHDIPTIPAGATNIEVKHNRQGRNNGSFLAIRAADGTYYILNGFTLSTLEQ 774
QY 537 DLVVKSLRYSGTGTAVESLQASRPILFTVEVLSVCKMPPRVRYSFYLPKPREDEK 596
Db 774 DLTYKGTVLRYSGSSAALERSFSLPEPLTIQVLMVGHALRPKIKFTYPMKKTES-- 832
QY 597 SSHPKDPRGSPVLSLHNSVLSNQVEQDDPPARWAGSWGPCSSASCGLQKRAVDRCG 656
Db 832 -----FNAIPTFS-----EWVIEWEGCSKTCGSGWQRRVYQCRD 867
QY 657 SAGQRTVPACDAH--RPVETOACGE-PCPTW 685
Db 867 INGH---PASECAKEVKPASTRPCADLPCPHW 896

```

RESULT 15
AAB21261
ID AAB21261 standard; Protein; 381 AA.
XX
AC AAB21261;
XX
DT 23-FEB-2001 (first entry)
XX
DE Human metalloproteinase ADAMTS-5.
XX
KW Human; ADAMTS-5; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; nootropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 196 /note= "Xaa= any amino acid"
FT Misc-difference 268 /note= "Xaa= any amino acid"
FT Misc-difference 304 /note= "Xaa= any amino acid"
FT Misc-difference 308 /note= "Xaa= any amino acid"
FT Misc-difference 308 /note= "Xaa= any amino acid"
XX
WO200053774-A2.
XX
PN 14-SEP-2000.
XX
PD 08-MAR-2000; 2000WO-US06237.
XX
PF 08-MAR-1999; 99US-0264585.
XX
PR (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
PA Kelner GS, Clark M, Maki RA;
XX
PI WPI; 2000-594326/56.
XX
DR N-PSDB; AAA95831.
XX
PT Polynucleotide encoding novel members of a disintegrin,
XX metalloproteinase and thrombospondin domain protein family used to
XX prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Claim 12; Fig 23; 129pp; English.
XX
CC The present sequence is human metalloproteinase ADAMTS-5. The
XX ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
XX and Metalloproteinase Domain) family. Members of the ADAMTS family
XX contain a thrombospondin domain in addition to the disintegrin and
XX metalloproteinase domains found in the ADAMTS. ADAMTS polypeptides are
XX useful for the manufacture of medicaments for treating conditions
XX associated with neuroinflammation and/or neurodegeneration, such as
XX Alzheimer's disease, Parkinson's disease and stroke. They are also
XX useful for treating conditions associated with cell proliferation, cell
XX migration, inflammation and/or angiogenesis, such as cancer, arthritis
XX and autoimmune diseases. They can be used to treat patients afflicted
XX with an invasive tumour, a brain tumour or brain injury.
XX
SQ Sequence 381 AA;

```

```

Query Match 48.6%; Score 1966; DB 21; Length 381;
Best Local Similarity 94.0%; Pred. No. 3.3e-147;
Matches 358; Conservative 3; Mismatches 18; Indels 2; Gaps 2;
XX
QY 28 HYLTLTAAALRYRHSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKLL 87
DB 2 HYRARAARAG-IFKHPHSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKLL 60

```

Search completed: April 29, 2003, 17:20:46
Job time : 45.9135 secs

```

QY 88 NKVSDKHPEYWDTAILFTRODLGCGATTCTDLGMADVGTMCMDPKRSCSVIEDDGLPSAFTT 147
DB 61 NKVSDKHPEYWDTAILFTRODLGCGATTCTDLGMADVGTMCMDPKRSCSVIEDDGLPSAFTT 120
QY 148 AHGLGHVFNPHDNVNVKCEEVFGKLRANHMSPTLQIDRANPWSACSAAIITDFLDLDSGH 207
DB 121 AHGLGHVFNPHDNVNVKCEEVFGKLRANHMSPTLQIDRANPWSACSAAIITDFLDLDSGH 180
QY 208 GDCLLDQPSKPISLPEDLPGASYTLSSQCELAFGVGSKPCPYMOCYCTKLWCTGKAKGMV 267
DB 181 GDCLLDQPSKPIFLPXDLPGASYTLSSQCELAFGVGSKPCPYMOCYCTKLWCTGKAKGMV 240
QY 268 CQTRHFPWADGTSCEGKGLKLGACVERHNLNKHVRVDSWAKWDPYGPCSRTCGGGVQLA 327
DB 241 CQTRHFPWADGTSCEGKGLKLGACVEXHNLNKHVRVDSWAKWDPYGPCSRTCGGGVQLA 300
QY 328 RROCTNPTP-ANGGKYCEGVVRYRSCNLEPCPSSASGKSFREOCEAFNGYNHSTNRLT 386
DB 301 RROXHQXPPLPTGGKYCEGVVRYRSCNLEPCPSSASGKSFREOCEAFNGYNHSTNRLT 360
QY 387 LAVAVVPKYSVGSVPRDKCKLI 407
DB 361 LAVAVVPKYSVGSVPRDKCKLI 381

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:25 ; Search time 46.2301 seconds
(without alignments)
2738.218 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 5164
Sequence: 1 MLLGILTLFAGRTAGGFE.....DQNLHRKPEQLDFCVLRPC 950

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5164	100.0	950	22	AAG62299 Human metalloprotease
2	5156	99.8	950	23	AAE22541 Human protease #2.
3	5111	99.0	952	23	AAU74751 Human protease PRT
4	4856.5	94.8	928	23	AAU72899 Human metalloprotease
5	2497.5	48.4	505	21	AAU72899 Rat metalloprotease
6	2480.5	48.0	950	20	AAU72899 Human METH1 protei
7	2480.5	48.0	950	22	AAU73549 Human ADAM-type me
8	2480.5	48.0	950	22	AAU73549 Human METH1. Homo
9	2480.5	48.0	967	19	AAU80285 Human integrin lig
10	2480.5	48.0	968	22	AAU50011 Protein; SEQ ID 12

11	2479.5	48.0	967	20	AAU04142 Human Tango-71 pro
12	2458.5	47.6	967	20	AAU78189 Human secreted pro
13	2445.5	47.4	950	21	AAU53899 Amino acid sequenc
14	2322	45.0	896	21	AAU21265 Mouse metalloprote
15	2274	44.0	727	20	AAU78435 Human ADAMTS-1 pro
16	2135.5	41.4	890	20	AAU49502 Human METH2 protei
17	2135.5	41.4	890	22	AAU50003 Human METH2. Homo
18	2126.5	41.2	889	22	AAU74946 Human ADAM type me
19	2113	40.9	905	22	AAU72284 Murine ADAMTS-8 am
20	1966	38.1	381	21	AAU21261 Human metalloprote
21	1965	38.1	1629	23	ABG30703 Human aggrecanase
22	1965	38.1	1629	23	AAU14448 Human ADAMTS-SI pr
23	1965	38.1	1916	23	AAU19173 Human protease, PR
24	1965	38.1	1935	23	AAU72896 Human metalloprote
25	1957	37.9	1602	23	ABG30702 Human aggrecanase
26	1943.5	37.6	1073	21	AAU21264 Human metalloprote
27	1939.5	37.6	1934	22	AAU72301 Human ADAMTS-9 alt
28	1917	37.1	837	20	AAU75425 Human aggrecan deg
29	1916	37.1	837	21	AAU99429 Human ADAMTS-9
30	1916	37.1	837	22	AAU29199 Human PRO1563 (UNQ
31	1916	37.1	837	22	AAU66178 Protein of the inv
32	1915	37.1	837	22	AAU78228 Human aggrecanase-
33	1915	37.1	840	21	AAU21256 Human metalloprote
34	1913	37.0	367	23	AAE22542 Human protease #3.
35	1903	36.9	680	21	AAU21251 Human metalloprote
36	1882	36.4	1907	23	AAU77133 Human protease #12
37	1874	36.3	1505	23	AAU72897 Human metalloprote
38	1872	36.3	947	22	AAU86950 Human metalloprote
39	1854	35.9	930	22	AAU72280 Murine ADAMTS-5 am
40	1848.5	35.8	1882	22	AAU72286 Human ADAMTS-9 am
41	1847.5	35.8	930	20	AAU75426 Human aggrecan deg
42	1840	35.6	929	21	AAU41226 Human ORFX ORF990
43	1798.5	34.8	874	22	AAU72287 Murine ADAMTS-9 am
44	1783	34.5	958	21	AAU21235 Human metalloprote
45	1762	34.1	870	21	AAU21252 Rat metalloprote

ALIGNMENTS

RESULT 1
AAG62299
ID AAG62299 standard; protein; 950 AA.
XX AC
XX AAG62299;
DT 23-AUG-2001 (first entry)
XX DT
XX Human metalloprotease MDT56 protein.
DE Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
KW osteopathic; antiarthritic.
XX KW
XX Homo sapiens.
OS
XX WO200134785-A1.
XX PD
XX 17-MAY-2001.
XX PF
XX 10-NOV-2000; 2000WO-JP07917.
XX PR
XX 11-NOV-1999; 99JP-0321740.
XX PR 16-MAY-2000; 2000JP-0144020.
XX PA (YAMA) YAMANOUCHI PHARM CO LTD.
XX PA (KAZU-) KAZUSA DNA RES INST.
XX PI Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX PI WPI; 2001-343602/36.
XX DR N-PSDB; AAH41003.
XX PT Metalloprotease with aggrecanase activity for treating joint diseases

PT especially osteoarthritis -
 PS Claim 1; Page 56-60; 85pp; Japanese.
 XX This invention relates to a metalloprotease with aggreganase activity.
 CC The invention includes protein and DNA sequences of the metalloprotease,
 CC vectors containing the DNA, host cells transformed by the vectors, and
 CC antibodies directed against the metalloprotease. The antibodies, protein
 CC and DNA sequences can be used in the treatment and prevention of joint
 CC diseases, particularly osteoarthritis. The treatment may result in
 CC osteopathic and antiarthritic activity. The present sequence represents
 CC the metalloprotease of the invention termed MDT56.
 XX
 SQ Sequence 950 AA;
 Query Match 100.0%; Score 5164; DB 22; Length 950;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 950; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 DB 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 QY 61 QEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGGSSDLRRCFCYSGDVNAEPDFAAVSLC 120
 DB 61 QEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGGSSDLRRCFCYSGDVNAEPDFAAVSLC 120
 QY 121 GGLRGAFYGAEEYVISPPLNASAPAAQNSQGAHLQRRGVPDSDPTSRGCVASGW 180
 DB 121 GGLRGAFYGAEEYVISPPLNASAPAAQNSQGAHLQRRGVPDSDPTSRGCVASGW 180
 QY 181 NPAILRALDPKPRAGGESSRRSRRAKRFVSIPIRYVETLVVADESVMVKEHADLEH 240
 DB 181 NPAILRALDPKPRAGGESSRRSRRAKRFVSIPIRYVETLVVADESVMVKEHADLEH 240
 QY 241 YLLTLATAARLYRHPSTLNPINIVVKVLLRRDSDGPKVTGNAALTFRNFCWQKLN 300
 DB 241 YLLTLATAARLYRHPSTLNPINIVVKVLLRRDSDGPKVTGNAALTFRNFCWQKLN 300
 QY 301 KVSQKHPEYWDATLFTQDLCGATTCDTLGMADVGMCDPKRSCVIEDDGLPSAFTTA 360
 DB 301 KVSQKHPEYWDATLFTQDLCGATTCDTLGMADVGMCDPKRSCVIEDDGLPSAFTTA 360
 QY 361 HELGHVFNPHNDNVKVEEVFGKLRANHMSPTLIQIDRANPWSACSAIIITDFLDSDGH 420
 DB 361 HELGHVFNPHNDNVKVEEVFGKLRANHMSPTLIQIDRANPWSACSAIIITDFLDSDGH 420
 QY 421 DCLLDQPSKPTSLPDLPGASYTLISQCELAFGVGSKEPCYMOYCTKLWCTGKAGQWVC 480
 DB 421 DCLLDQPSKPTSLPDLPGASYTLISQCELAFGVGSKEPCYMOYCTKLWCTGKAGQWVC 480
 QY 481 QTRHFPWADGTSCGSGKCLKAGACVERHNLNKRVGDSWAKWDYPCGSRCTCGGVQLAR 540
 DB 481 QTRHFPWADGTSCGSGKCLKAGACVERHNLNKRVGDSWAKWDYPCGSRCTCGGVQLAR 540
 QY 541 RQCTNPTPANGKYCEGVVRYRSCNLEPCPSASAGKSFREOCEAFNGYHNSTNRLTLA 600
 DB 541 RQCTNPTPANGKYCEGVVRYRSCNLEPCPSASAGKSFREOCEAFNGYHNSTNRLTLA 600
 QY 601 VAWVPKYSVSPRDKKLCIRANGTYFYVLAPKVVDGTLCSPDSTSVCVGKCIKAGCD 660
 DB 601 VAWVPKYSVSPRDKKLCIRANGTYFYVLAPKVVDGTLCSPDSTSVCVGKCIKAGCD 660
 QY 661 GNLGSKKRPDKCGVCGDNKSKCKVTGLTKPMHGYNFVVAIPAGASSIDIRQRYKGLI 720
 DB 661 GNLGSKKRPDKCGVCGDNKSKCKVTGLTKPMHGYNFVVAIPAGASSIDIRQRYKGLI 720
 QY 721 GDNVYALKNSOGKYLNLGHFVSAVERDLVVKGSLLRYSGTGAVESLQASRPILPILT 780
 DB 721 GDNVYALKNSOGKYLNLGHFVSAVERDLVVKGSLLRYSGTGAVESLQASRPILPILT 780
 QY 781 VEVLVSGKMTPPRVRYSFYLPKEPREDKSSHFKDPRGFSVLHNSVLSLNOVEQPDDRPP 840

DB 781 VEVLVSGKMTPPRVRYSFYLPKEPREDKSSHFKDPRGFSVLHNSVLSLNOVEQPDDRPP 840
 QY 841 ARWVAGSWGPCASCSGSLQKRAVDCRGSAQORTVPACDAARHPVETQACGECPTWELS 900
 DB 841 ARWVAGSWGPCASCSGSLQKRAVDCRGSAQORTVPACDAARHPVETQACGECPTWELS 900
 QY 901 ANSPCKSKCGRGFORSLKCVGHGRLRLARDQCNLHRRKPOELDFCVLRPC 950
 DB 901 ANSPCKSKCGRGFORSLKCVGHGRLRLARDQCNLHRRKPOELDFCVLRPC 950
 RESULT 2
 AAE22541
 ID AAE22541 standard; Protein; 950 AA.
 XX
 AC AAE22541;
 XX
 DT 26-JUL-2002 (first entry)
 XX
 DE Human protease #2.
 KW Human; novel human protein; NHP; protease; biological disorder; obesity;
 KW high blood pressure; arthritis; connective tissue disorder; infertility;
 KW gene therapy; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN WO200225949-A2.
 PD 04-APR-2002.
 XX
 PF 27-SEP-2001; 2001WO-US30350.
 XX
 PR 29-SEP-2000; 2000US-236689P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Friddle CJ, Hilbun E;
 DR WPI; 2002-372123/40.
 DR N-PSDB; AAD35569.
 XX
 PT Novel nucleic acid encoding a human protease, useful as a hybridization
 PT probe for screening libraries and assessing gene expression patterns -
 XX
 PS Claim 6; Page 36-38; 4lpp; English.
 XX
 CC The present sequence is novel human protein (NHP), human protease.
 CC NHPs share structural similarity with animal proteases particularly
 CC zinc metalloproteases. Sequences of the invention are useful in
 CC therapeutic, diagnostic and pharmacogenomic applications. NHP
 CC polynucleotides are used as hybridisation probes for screening
 CC libraries and assessing gene expression patterns. They can also be
 CC used for treating related biological disorders such as obesity, high
 CC blood pressure, arthritis, connective tissue disorders and infertility.
 CC They are also used in gene therapy.
 XX
 SQ Sequence 950 AA;
 Query Match 99.8%; Score 5156; DB 23; Length 950;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 949; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 DB 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 QY 61 QEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGGSSDLRRCFCYSGDVNAEPDFAAVSLC 120
 DB 61 QEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGGSSDLRRCFCYSGDVNAEPDFAAVSLC 120
 QY 121 GGLRGAFYGAEEYVISPPLNASAPAAQNSQGAHLQRRGVPDSDPTSRGCVASGW 180

Db 121 GGLRGAFGYRGAEYVISLPNAPSAPAAQNSOGAHLQRRGVPGSPGDPTSRCGVASGW 180
QY 181 NPAILRALDPYKPRAGCESRRSRSGRAKRFVSIPIRYVETLVVADESWMKFGADLEH 240
Db 181 NPAILRALDPYKPRAGCESRRSRSGRAKRFVSIPIRYVETLVVADESWMKFGADLEH 240
QY 241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRSDGPKVTGNAALTLRNFCAWOKKLN 300
Db 241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRSDGPKVTGNAALTLRNFCAWOKKLN 300
QY 301 KYSDKHPEYWDTAIFLTRDLCGATCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA 360
Db 301 KYSDKHPEYWDTAIFLTRDLCGATCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA 360
QY 361 HELGHVFNPHDNVAVCEVFGKLRANHMMSPTLIQIDRANPWSACSAAIITDFLDSHG 420
Db 361 HELGHVFNPHDNVAVCEVFGKLRANHMMSPTLIQIDRANPWSACSAAIITDFLDSHG 420
QY 421 DCLLDQPSKPISLPDLPGASYTLSSQCELAFAFGVSGKPCPYMQYCTKLWCTGKAKQWVC 480
Db 421 DCLLDQPSKPISLPDLPGASYTLSSQCELAFAFGVSGKPCPYMQYCTKLWCTGKAKQWVC 480
QY 481 QTRHFPWADTSCGSEKLCAGACVERHNLNKHRVDGSAKWDPYGPCSRTCGGVQLAR 540
Db 481 QTRHFPWADTSCGSEKLCAGACVERHNLNKHRVDGSAKWDPYGPCSRTCGGVQLAR 540
QY 541 RQCTNPTPANGKGYCEGVKRVKSCNLEPCPSSASGKSFREQCEAFNGYNHSTNRLTLA 600
Db 541 RQCTNPTPANGKGYCEGVKRVKSCNLEPCPSSASGKSFREQCEAFNGYNHSTNRLTLA 600
QY 601 VAWPKYSGVSPDRCKKLCIRANGTGYFYVLAPKYVVDGTLGSPDSTSVCGVKCIKAGCD 660
Db 601 VAWPKYSGVSPDRCKKLCIRANGTGYFYVLAPKYVVDGTLGSPDSTSVCGVKCIKAGCD 660
QY 661 GNLSGKRRDKCGVGGDNKCKYTGFTKPMHGYNFWAIPAGASSIDIRQGYKGLI 720
Db 661 GNLSGKRRDKCGVGGDNKCKYTGFTKPMHGYNFWAIPAGASSIDIRQGYKGLI 720
QY 721 GDNYLAKNSGKYLNGHGFVYSAVERDLVVKGLLSYTGTTAVESLQASRPILPLEPT 780
Db 721 GDNYLAKNSGKYLNGHGFVYSAVERDLVVKGLLSYTGTTAVESLQASRPILPLEPT 780
QY 781 VEVLVSGKMTPRVRYSYFLPKPEPREDKSHPKDPRGVSVLHNSVLSNQVEQDDRRPP 840
Db 781 VEVLVSGKMTPRVRYSYFLPKPEPREDKSHPKDPRGVSVLHNSVLSNQVEQDDRRPP 840
QY 841 ARWVAGSWGPCASGSGGLQKRAVDCRSAGORTVPACDAARHPVETQACGEPCTWELS 900
Db 841 ARWVAGSWGPCASGSGGLQKRAVDCRSAGORTVPACDAARHPVETQACGEPCTWELS 900
QY 901 AWSPCSKSGRGFRRLSKCVGHGGLRLARDQCNLHRKPQELDFCVLRPC 950
Db 901 AWSPCSKSGRGFRRLSKCVGHGGLRLARDQCNLHRKPQELDFCVLRPC 950
RESULT 3
ID AAU74751
XX- AAU74751 standard; Protein; 952 AA.
AC AAU74751;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-11 protein sequence.
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis.
OS Homo sapiens.

XX WO200198468-A2.
PN
XX
PD 27-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-US19178.
XX
PR 16-JUN-2000; 2000US-2123336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
PR 14-JUL-2000; 2000US-218946P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Deleagane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Waleia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX
DR WPI; 2002-090437/12.
DR N-PSDB; ABK12894.
XX
XX
PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX
PS Claim 1; Page 144-146; 177pp; English.
XX
CC The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present protein sequence represents
CC the human protease PRTS-11 protein of the invention.
XX
SQ Sequence 952 AA;
Query Match 99.0%; Score 5111; DB 23; Length 952;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 946; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
QY 1 MLLGLITLAFAGTAGGEPEREVVPVPIRLDPDINGRRYRWGPDGGLIFQITAF 60
Db 1 MLLGLITLAFAGTAGGEPEREVVPVPIRLDPDINGRRYRWGPDGGLIFQITAF 60
QY 61 QEDFYHLTDAQFLAPAFSTEHLGVLPQGLTGSSDLRRCFYSGDVNAEPDFAVSLC 120
Db 61 QEDFYHLTDAQFLAPAFSTEHLGVLPQGLTGSSDLRRCFYSGDVNAEPDFAVSLC 120
QY 121 GGLRGAFGYRGAEYVISLPNAPSAPAAQNSOGAHLQRRGVPGSPGDPTSRCGVASGW 180
Db 121 GGLRGAFGYRGAEYVISLPNAPSAPAAQNSOGAHLQRRGVPGSPGDPTSRCGVASGW 180
QY 181 NPAILRALDPYKPRAGCESRRSRSGRAKRFVSIPIRYVETLVVADESWMKFGADLEH 240
Db 181 NPAILRALDPYKPRAGCESRRSRSGRAKRFVSIPIRYVETLVVADESWMKFGADLEH 240
QY 241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRSDGPKVTGNAALTLRNFCAWOKKLN 300
Db 241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRSDGPKVTGNAALTLRNFCAWOKKLN 300
QY 301 KYSDKHPEYWDTAIFLTRDLCGATCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA 360
Db 301 KYSDKHPEYWDTAIFLTRDLCGATCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA 360

Db 301 KVSQKHEPYWDTAILFTRODLGATTCDTLGMADVMTCDPKRSCSVIEDGLPSAFTTA 360
 QY 361 HELGHVFNMPHDNVKCEEEVFGKLRANHMSPPTLIQIDRANPWSACSAAITDLDLSHG 420
 Db 361 HELGHVFNMPHDNVKCEEEVFGKLRANHMSPPTLIQIDRANPWSACSAAITDLDLSHG 420
 QY 421 DCLLDQSPKPSLPELPGASYTSLQOCELAFCVGVSKPCPYMOYCTKLMCTGKAKQWVC 480
 Db 421 DCLLDQSPKPSLPELPGASYTSLQOCELAFCVGVSKPCPYMOYCTKLMCTGKAKQWVC 480
 QY 481 QTRHFPWADTSCGEGKLCIKGACVERHNLKRVNKGWAKWDPYPCSRCTGGGVQLAR 540
 Db 481 QTRHFPWADTSCGEGKLCIKGACVERHNLKRVNKGWAKWDPYPCSRCTGGGVQLAR 540
 QY 541 RQCTNPTPANGKYCEGVRYKYSCLNLEPCPSASCKSPREOCEAFNGYNHSTNRLTLA 600
 Db 541 RQCTNPTPANGKYCEGVRYKYSCLNLEPCPSASCKSPREOCEAFNGYNHSTNRLTLA 600
 QY 601 VAWVPKYSGVSPRDKCLICRANGTYFYVYLAPK - VVDGTLCSPDSTSVCVQKCIKAGC 659
 Db 601 VAWVPKYSGVSPRDKCLICRANGTYFYVYLAPKVVVDGTLCSPDSTSVCVQKCIKAGC 660
 QY 660 DGNLGSKKRDKCGVCGGDNKCKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQYKGL 719
 Db 661 DGNLGSKKRDKCGVCGGDNKCKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQYKGL 720
 QY 720 IGDNYLALANSQKYLNLNGHFVYSAVERDLVVKGSLLRYSGTGAVESLQASRPILPL 779
 Db 721 IGDNYLALANSQKYLNLNGHFVYSAVERDLVVKGSLLRYSGTGAVESLQASRPILPL 780
 QY 780 TVEVLSVGKMTPPRVRYSFYLPKEPREDKSSHPKDR - GPSVLHNSVLSNQVEQPDOR 838
 Db 781 TVEVLSVGKMTPPRVRYSFYLPKEPREDKSSHPPHPRGGPSVLHNSVLSNQVEQPDOR 840
 QY 839 PPARWAGSGPCASGSGQLQRAVDCRGSGQRTVPACDAHRPVEVQACGECPTWE 898
 Db 841 PPARWAGSGPCASGSGQLQRAVDCRGSGQRTVPACDAHRPVEVQACGECPTWE 900
 QY 899 LSWSPCSKSCGRGFORRLSKVGHGRLRLARDQCNLHRRKPQLDPCVLRPC 950
 Db 901 LSWSPCSKSCGRGFORRLSKVGHGRLRLARDQCNLHRRKPQLDPCVLRPC 952

RESULT 4
 AAU72899
 ID AAU72899 standard; Protein: 928 AA.
 XX AC AAU72899;
 XX AC AAU72899;
 XX DT 26-FEB-2002 (first entry)
 XX Human metalloprotease partial protein sequence #11.
 XX Human; protease; PCR primer; cytostatic; immunomodulator; cardiac;
 KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
 KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
 KW anorectic; antinflammatory; aspartyl protease; cysteine protease;
 KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;
 KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
 KW immune-related disease; cardiovascular disease; neuronal disease;
 KW migraine; sexual dysfunction; mood disorder; attention disorder;
 KW cognition disorder; hypotension; hypertension; psychotic disorder;
 *KW dyskinesia; metabolic disorder; inflammatory disorder.
 XX OS Homo sapiens.
 XX OS W0200183782-A2.
 XX PN 08-NOV-2001.
 XX PF 04-MAY-2001; 2001WO-US14431.
 XX PF 04-MAY-2000; 2000US-201879P.

(SUGE-) SUGEN INC.
 Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 Payne V;
 WPI; 2002-041502/05.
 DR N-PSDB; AAS97182.
 XX Novel protease polypeptide useful for screening for substances that may
 be used to treat, e.g., cancers, immune-related diseases,
 PT cardiovascular disease, migraine, pain, psychotic and inflammatory
 PT disorders -
 XX Claim 28; Figure 2G; 232pp; English.
 XX The invention relates to an isolated, enriched, or purified protease
 CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
 CC screen for substances (S) that may modulate its activity. Administering
 CC S (which modulates protease activity in vitro) may be used to treat a
 CC disease or disorder selected from cancers (e.g., of tissues, of blood or
 CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
 CC brain, ovarian, bladder or kidney), immune-related diseases and
 CC disorders, cardiovascular disease, brain or neuronal-associated diseases
 CC (e.g., central or peripheral nervous system diseases, migraine, pain,
 CC sexual dysfunction, mood disorders, attention disorders, cognition
 CC disorders, hypotension, hypertension, psychotic disorders, neurological
 CC disorders and dyskinesias), metabolic disorders and inflammatory
 CC disorders. (I) may also be useful as a diagnostic tool for a disease or
 CC disorder such as those above. AAU72876-AAU72910 represent human
 CC protease amino acid sequences of the invention.
 XX SQ Sequence 928 AA;
 Query Match 94.0%; Score 4856.5; DB 23; Length 928;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 901; Conservative 0; Mismatches 1; Indels 49; Gaps 2;
 QY 1 MLLLGILTLAFAGTAGGFEPEYVVPVIRLDPDINGRRYVWRGPDGGLQIFQITAF 60
 Db 26 MLLLGILTLAFAGTAGGFEPEYVVPVIRLDPDINGRRYVWRGPDGGLQIFQITAF 85
 QY 61 QEDFYLHITPAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDFAAAYSLC 120
 Db 86 QEDFYLHITPAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDFAAAYSLC 145
 QY 121 GGLGAFGYGAEYVISPLPNASAPAAQNSQGAHLQRRGVPGGSDPTSRGCVASGW 180
 Db 146 GGLGAFGYGAEYVISPLPNASAPAAQNSQGAHLQRRGVPGGSDPTSRGCVASGW 205
 QY 181 NPAILRALDPYKRRAGFGESRRRRSRRAKRFVSIPIRYVETLVVVADESVMKFGADLEH 240
 Db 206 NPAILRALDPYKRRAGFGESRRRRSRRAKRFVSIPIRYVETLVVVADESVMKFGADLEH 265
 QY 241 YLLTLTAARLYRHPSTILNINIVVXVLLLRDSDGPKVTGNALTLRNFCAWQKLN 300
 Db 266 YLLTLTAARLYRHPSTILNINIVVXVLLLRDSDGPKVTGNALTLRNFCAWQKLN 325
 QY 301 KVSQKHEPYWDTAILFTRODLGATTCDTLGMADVMTCDPKRSCSVIEDGLPSAFTTA 360
 Db 326 KVSQKHEPYWDTAILFTRODLGATTCDTLGMADVMTCDPKRSCSVIEDGLPSAFTTA 385
 QY 361 HELGHVFNMPHDNVKCEEEVFGKLRANHMSPPTLIQIDRANPWSACSAAITDLDLSHG 420
 Db 386 HELGHVFNMPHDNVKCEEEVFGKLRANHMSPPTLIQIDRANPWSACSAAITDLDLSHG 445
 QY 421 DCLLDQSPKPSLPELPGASYTSLQOCELAFCVGVSKPCPYMOYCTKLMCTGKAKQWVC 480
 Db 446 DCLLDQSPKPSLPELPGASYTSLQOCELAFCVGVSKPCPYMOYCTKLMCTGKAKQWVC 505
 QY 481 QTRHFPWADTSCGEGKLCIKGACVERHNLKRVNKGWAKWDPYPCSRCTGGGVQLAR 540
 Db 506 QTRHFPWADTSCGEGKLCIKGACVERHNLKRVNKGWAKWDPYPCSRCTGGGVQLAR 565

QY 541 ROCTNPTPANGKYCEGVKRYKNSCNLEPCPSSASGKSFREOCEAFNGYNHSTNRLTLA 600
 DB 566 ROCTNPTPANGKYCEGVKRYKNSCNLEPCPSSASGKSFREOCEAFNGYNHSTNRLTLA 625
 QY 601 VAWPKYSGVSPDKCKLICRANGTGYFYVLAPK-VVDGTLCSDDSTSVQVQKCIKAGC 659
 DB 626 VAWPKYSGVSPDKCKLICRANGTGYFYVLAPKVVVDGTLCSDDSTSVQVQKCIKAGC 685
 QY 660 DGNLGSKKRFRDCGVCGGDNKCKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQGYKGL 719
 DB 686 DGNLGSKKRFRDCGVCGGDNKCKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQGYKGL 745
 QY 720 IGDNDYLAIKNSGKYLLNGHFFVSAVERDLVVGSLRLRYSGTGTAVESLQASRPILLEPL 779
 DB 746 IGDNDYLAIKNSGKYLLNGHFFVSAVERDLVVGSLRLRYSGTGTAVESLQASRPILLEPL 800
 QY 780 TVEVLSVGKWTTPRVRYSPFLPKPEPKEDKSHPKDPRGSPVLSLNSQVQEQDDRP 839
 DB 801 TVEVLSVGKWTTPRVRYSPFLPKPEPKEDKSHPKDPRGSPVLSLNSQVQEQDDRP 817
 QY 840 PARWAGSGPCASGSGLOKRAVDCRGSGAGQRTVPACDAHRPVETQACGECPTWEL 899
 DB 818 PARWAGSGPCASGSGLOKRAVDCRGSGAGQRTVPACDAHRPVETQACGECPTWEL 877
 QY 900 SAWSPCKSGRGFRORSILKCVGHGRLRLARDQCNLHRKPOELDFCVLRPC 950
 DB 878 SAWSPCKSGRGFRORSILKCVGHGRLRLARDQCNLHRKPOELDFCVLRPC 928
 RESULT 5
 AAB21257
 ID AAB21257 standard; Protein; 505 AA.
 XX
 AC AAB21257;
 XX
 XX 23-FEB-2001 (first entry)
 XX
 DE Rat metalloproteinase ADAMTS-5.
 XX
 KW Rat; ADAMTS-5; metalloproteinase; ADAM;
 KW a disintegrin and metalloproteinase domain; thrombospondin domain;
 KW vaccine; neurotrophic; neuroprotective; antiparkinsonian;
 KW cerebroprotective; cytotatic; antiarthritic; immunosuppressive;
 KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
 KW autoimmune disease; brain tumour; brain injury.
 XX
 OS Rattus norvegicus.
 XX
 PN WO200053774-A2.
 XX
 XX 14-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US06237.
 XX
 PR 08-MAR-1999; 99US-0264585.
 XX
 PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
 XX
 PI Keiner GS, Clark M, Maki RA;
 XX
 XX WPI: 2000-594326/56.
 DR N-PSDB; AAN95827.
 XX
 XX Polynucleotide encoding novel members of a disintegrin,
 PT metalloproteinase and thrombospondin domain protein family used to
 PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
 XX
 XX Claim 12; Fig 14; 129pp; English.
 PS
 XX The present sequence is rat metalloproteinase ADAMTS-5. The
 CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
 CC and Metalloproteinase Domain) family. Members of the ADAMTS family

CC contain a thrombospondin domain in addition to the disintegrin and
 CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
 CC useful for the manufacture of medicaments for treating conditions
 CC associated with neuroinflammation and/or neurodegeneration, such as
 CC Alzheimer's disease, Parkinson's disease and stroke. They are also
 CC useful for treating conditions associated with cell proliferation, cell
 CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
 CC and autoimmune diseases. They can be used to treat patients afflicted
 CC with an invasive tumour, a brain tumour or brain injury.
 XX
 SQ Sequence 505 AA:
 Query Match 48.4%; Score 2497.5; DB 21; Length 505;
 Best Local Similarity 93.5%; Pred. No. 1.3e-185;
 Matches 462; Conservative 8; Mismatches 21; Indels 3; Gaps 2;
 QY 89 QGLTGSDDLRCRCFYSGDVNAEPDPSFAAVSLCGGLRGAFGYGAEYVISPLNAPAAQ 148
 DB 13 QRLTGSDDLRCRCFYSGDVNAEPDPSFAAVSLCGGLRGAFGYGAEYVISPLNAPAAQ 72
 QY 149 RNSQGAHLQRRGVPGSGDPTSRGVSAGWNPAIRALDPYKPRRAGFGESRRRSRG 208
 DB 73 RNSQGAHLQRRGVPGSGDPTSRGVSAGWNPAIRALDPYKPRRAGFGESRRRSRG 132
 QY 209 RAKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLAARLYRHPISILNINIVVK 268
 DB 133 RAKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLAARLYRHPISILNINIVVK 192
 QY 269 VLLERDRSGPKVTGNAALTNRNFCAMOKKLNKVSFKHPEYWDTAILFTRODLGATTC 328
 DB 193 VLLERDRSGPKVTGNAALTNRNFCAMOKKLNKVSFKHPEYWDTAILFTRODLGATTC 252
 QY 329 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHGLGHVFNPHDNVKEEVFGKLRANH 388
 DB 253 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHGLGHVFNPHDNVKEEVFGKLRANH 312
 QY 389 MMSPTLIQIDRANPWSACSAATITDLSGHGDCLLDQSPKISLPEDLPGASTYLSQOC 448
 DB 313 MMSPTLIQIDRANPWSACSAATITDLSGHGDCLLDQSPKISLPEDLPGASTYLSQOC 372
 QY 449 ELAFGVSGPCPYMOCYCTKLWCTGKAKGQWCVOTRHFPAWDTSCGEGKLCIGACVERH 508
 DB 373 ELAFGVSGPCPYMOCYCTKLWCTGKAKGQWCVOTRHFPAWDTSCGEGKLCIGACVERH 432
 QY 509 NLNKHVDSGSAKWDYPGCSRTCCGGVQLARR--QCTNPTPANGKYCEGVRYKRSN 566
 DB 433 NPKNYRVDGPAKWEYPGCSRTCCGGVQLARR--QCTNPTPANGKYCEGVRYKRSN 491
 QY 567 LEPCPSSASGKSF 580
 DB 492 LEPCPSSASGKSF 505
 RESULT 6
 AAY49501
 ID AAY49501 standard; Protein; 950 AA.
 XX
 AC AAY49501;
 XX
 XX 10-JAN-2000 (first entry)
 XX
 DE Human METH1 protein.
 XX
 XX Human; METH1; MEH2; anti-angiogenic; metalloproteinase thrombospondin;
 KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
 KW angiogenesis inhibitor; abnormal wound healing; inflammation;
 KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
 KW diabetic retinopathy; macula degeneration; haemangioma; detection;
 KW arterial-venous malformation; immune deficiency.
 XX
 OS Homo sapiens.
 XX
 PN WO9937660-A1.

XX PD 29-JUL-1999. 99MO-US01313.

XX PF 22-JAN-1999; 98US-0072298.

XX PR 23-JAN-1998; 98US-0098539.

XX PR 28-AUG-1998; 98US-0098539.

XX PA (IRUE/) IRUELA-ARISPE L.

XX PA (HAST/) HASTINGS G A.

XX PA (RUBE/) RUBEN S M.

XX PI IrueLa-Arispe L, Hastings GA, Ruben SM;

XX PI WPI; 1999-590684/50.

XX DR N-PSDB; AA232000.

XX PT New isolated metalloprotease thrombospondin polypeptides, useful for

XX PT treating hyperproliferative disorders, cancers or autoimmune disorders

XX PS Claim 10; Fig 1; 457pp; English.

XX AA232000 and AA232001 encode, and AA49501 and AA49502 represent, human

XX metalloprotease thrombospondin (METH) proteins METH1 and METH2

XX respectively. METH1 and METH2 have been found to be potent inhibitors of

XX angiogenesis both in vitro and in vivo. They can be used for treating

XX cancer and other disorders related to angiogenesis including abnormal

XX wound healing, inflammation, rheumatoid arthritis, psoriasis,

XX endometrial bleeding disorders, diabetic retinopathy, some forms of

XX macula degeneration, haemangiomas, and arterial-venous malformations.

XX They may be useful in treating deficiencies or disorders of the immune

XX system, by activating or inhibiting the proliferation, differentiation,

XX or mobilization (chemotaxis) of immune cells. The etiology of these

XX immune deficiencies or disorders may be genetic, somatic, such as

XX cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or

XX toxins), or infectious. They can also be used to treat inflammatory

XX conditions, both chronic and acute conditions. The products can also be

XX used for detection and diagnosis. AA232002 to AA232080, and AA49503 to

XX AA49511 represent sequences given in the exemplification of the present

XX invention.

XX Sequence 950 AA;

Query Match 48.0%; Score 2480.5; DB 20; Length 950;

Best Local Similarity 48.9%; Pred. No. 6.8e-184;

Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY 1 MLLGLILFLAFAGTAGGFEPEVEVVPRLDPDINGRRYYWRGPEDSGDOGLIFQITAF 60

DB 19 LLLAAALLAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTTLRRLHAF 64

QY 61 QEDFYLHLPDPAFLAPAFSTHGLG-----VPLQGLTGGSSDLRRCFYSGDVNAEPDSF 114

DB 65 DQDLDELRLPSSFLAPGFTIONVGRKSGSETPLP-----ETDLACHFYSGTVNGDPSSA 119

QY 115 AAVSLCGGLRGAFYRGAEYVISP LNAS---APAAQNSOGA-----HLLQ--RRGVPDG 165

DB 120 AALSCEGVRGAFYLLGEAYFTQPLPAASERLATAAPGKPPAPLOQFHLRRNRQGDYVG 179

QY 166 PSG-----DPTSR-----GVASG--WNPAILRALDPYKPRRAGRSRRRS 207

DB 180 TCGVDDDEPRPTGKAETEDDEGTGEDEGPQMS-----PQDPALQGVQCP--TGTS 230

QY 208 GRAKREVSIPRVVETLVVADESNVAFHGADLEHYLLTLATAARLYRHPSTLINPIVYV 267

DB 231 IRKKREVSRRVETMLVADQSMAEFHGSLGKHYLLTLFSVAARLYKHPSTIRNSVLVV 290

QY 268 KVLRLDRSDGPKVTGNAALTIRNFCAMOKLNKVDKHPYWDTAIFLTRODLGATTC 327

DB 291 KILVHDEQKQPEVTSNAALTIRNFCNMOKHNPSPDRDAEHYDTAIFLTRODLGSGT 350

QY 328 DTLGMDVGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNDVNVCEVFCKLKRN 387

DB 351 DTLGMDVGTVCDPKSCSVIEDDGLQAAFTTAHELGHVFNPHDNDVNVCEVFCKLKRN 410

QY 388 HMMSTLTIQIDRANPWSACSRAIITDFLDSDGDCLLDQPSKPSLSPDLPGASVYTL 447

DB 411 HMMASMLNLDHSQWSPSCSAYMITSFJLDNGHGECMLDKPONPQLPGDLFGTSDANRQ 470

QY 448 CELAFGVGSKPCP-YMQYCTKLWCTGKAKGOMVCOTRHPWADGTSCGEGKLCUKGACVE 506

DB 471 CQFTFGEDSKHCPDAASTCSLWCTGTSGGLVLCQTKHFPWADGTSCGEGKWCINGKCVN 530

QY 507 RHNLNKH---RVDGSKWAKWDYPGCSRTCCGGVQLARQCTNPTPANGKCYCEGVVKYR 563

DB 531 KTD-RKHEDTPPHGSMGMPGDCSRTCCGGVQYTMRECONPVKNGKCYCEGKRVYR 589

QY 564 SCNLEPCSSASGSKSFRQCEAFNGYNHSTNRLTLAVAWVPKYSGVSPRDKCKLICRAN 623

DB 590 SCNLEPCPN-NKTFREQCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLICQAK 648

QY 624 GTGYFYVLAPKVVDTGLCSPDSTSVCVQKCKIACGDNGLSKKRPDKGCGVCGGDNKSK 683

DB 649 GIGYFFVLQPKVVDTGTPCSPDSTSVCVQKCKIACGDRIDSKKFKDKCGVCGGNGSTCK 708

QY 684 KVTGLETKPMHGYNFVVAIPAGASSIDIRQGYKGLIGDNDVNLKNSOGKYLNGHFV 743

DB 709 KISGVSITSAKPGYHDITITPTGATNIEVAKQNRGSRNNGSFLAIKAADGTVILNGDYTL 768

QY 744 SAVERDLVWKGSLRYSGTGATVESLQASRPIDLEPLTVEVLSVGMKTPPRVRYSPYLPKE 803

DB 769 STLEQDIHYKGVLYRYSGSSAALERISFSLKEPLTIQVLTGNALRPKIKYTFVAKK 828

QY 804 PREKSSHPDKRPGSVLNSVLNSQVEQDDPPAPBFWAGSWGSCSASCSGSLQKRA 863

DB 829 ---KES-----FNAIPTFS-----AWTIEEWGECSEKCELAGWQRL 861

QY 864 VDCGSGAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCKSKSGRGFORSLK 920

DB 862 VECDINGQ---PASECAKEVPASTPCADHPCPQWOLGSEWSSCKTCGKGYKRRSLK 918

QY 921 VHGGLLRLARDQCNLHKKPQE-LDFCVLRPC 950

DB 919 LSHDGGVLSHESCDPLKPKHFIDFCTMAEC 949

RESULT 7

AA232000

ID AAB73549 standard; Protein; 950 AA.

XX AAB73549;

XX AC AAB73549;

XX DT 07-AUG-2001 (first entry)

XX DE Human ADAM-type metalloprotease MDT54, SEQ ID NO:4.

XX KW Human; MDT54; ADAM-type metalloprotease; drug screening;

XX KW A Disintegrin And Metalloprotease; cancer; arthritis.

XX OS Homo sapiens.

XX PN JP2001017183-A.

XX PD 23-JAN-2001.

XX PF 09-JUL-1999; 99JP-0196584.

XX PR 09-JUL-1999; 99JP-0196584.

XX PA (YAMA) YAMANOUCHI PHARM CO LTD.

XX WPI; 2001-275950/29.

XX DR N-PSDB; AAH20224.

XX PT A new metal protease and its preparation for use as an anti-cancer and

PT anti-arthritis therapeutic -

PS Claim 1; Page 12-14; 22pp; Japanese.

CC The invention relates to the novel human ADAM (A Disintegrin And
CC Metalloprotease)-type metalloproteases MDTs4 (AAB/3549) and MDTs5
CC (AAB/3550). The metalloproteases can be used for the treatment of
CC cancers and arthritis. The invention also relates to the genes encoding
CC MDTs4 and MDTs5, vectors and host cells containing the MDTs4 or MDTs5
CC genes, the recombinant production of MDTs4 and MDTs5, and antibody
CC specific for MDTs4 or MDTs5, and methods of screening for compounds
CC which modulate the activity of MDTs4 and/or MDTs5. The present
CC sequence represents human MDTs4.

SX Sequence 950 AA;

Query Match 48.08; Score 2480.5; DB 22; Length 950;
Best Local Similarity 48.99; Pred. No. 6.8e-184;
Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY 1 MLLGILTLAPAGTAGGFEPEVEVPIRLDPDINGRYYWRGPGSDGQGLIFQITAF 60
DB 19 LLLAAALLAVSDALGRSEDEELVP-ELE-----RAP---GHGTTLRLLHAF 64
QY 61 QEDFYLHTLTPAQFLAPAFSTEHLG-----VPLQGLTGSSDLRRCFYSGDVNAEPDSF 114
DB 65 DQDLLELRPDSSFLAPGFTLQNGVRKSGSETPLP-----ETDLAHCFYSTVNGDPSSA 119
QY 115 AAVSLCGGLGAFYGRGAEYISLPLNAS---APAAQNSOGA---HLIQ--RRGVPGG 165
DB 120 AALSCEGVRGAYLLGEAYFIQPLPAASERLATAAPEGKPPAPLOPHLLRRNRQGVGG 179
QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKRRAGFSGRSRRS 207
DB 180 TCGVVDDPRPTGKAETDEDEGEDEGPQWS-----PDPALQGVGP-TCGTS 230
QY 208 GRAKRFVSIPIRYETLVVADSNVKGFGADLEHYLLILLATAAALYRHPSTLNPINIVV 267
DB 231 IRKKRFVSSHRYETMLVADQSMAEFGHGLKHYLLTLFSAARLYRHPSTIRNSVLVV 290
QY 268 KVLILRDSDGPKVTGNAALTLRFCAWQKLNKVSQKHPEYWDTAILFTQDLGATTC 327
DB 291 KILVHDEQKGPVTSNAALTIRNFCWQKHNPSPDRDAEHYDTALFTQDLGSGQTC 350
QY 328 DTLMADVMTWCDPKRSVIEDDGLPSAFTHAHELGHVFNMPHDNVKCEVEFGKLRAN 387
DB 351 DTLMADVMTWCDPKRSVIEDDGLQAAFTTAHELGHVFNMPHDNAKQCASLNGVQDS 410
QY 388 HWSPTLIQIDRANPWSACSAIITDPLDSHGDCILDDQSKPLSLPEDLPGLASVYLSQQ 447
DB 411 HMASMLNSLDHSPFSPCSAYMITTSFLDNGHGECMLDKFQNPQLPGLDLPSTYDANRQ 470
QY 448 CELAFGVGSKPCP-YMOYCTKLWCTGKAKGMVQOTRHFPAWDTSCGEGKLCILKACVE 506
DB 471 CQFTFGEDSKHCPDAATCSTLWCTGTSGGVLCQTHFFWADGTSCEGKWCINGKCVN 530
QY 507 RHNLNK---RVDGSAKWPYGPSCSTCGGVLQARRQCTNPTPANGKYCEGVRVYR 563
DB 531 KTD-RKHFDTPFHGSGWGMFGWDCSRTCGGVQVYTMRECDNPVKNKGKYCEGVRVYR 589
QY 564 SCNLEPCPSSAGSKSFREOCEAFNGYNHSTNRLTAVAMVPKYSVGSVPDKCKLICRAN 623
DB 590 SCNLEDCPDN-NGKTFREOCEAFNEFSKASFGSGPAVEWIPKYAVGSPDKRCKLICQAK 648
QY 624 GTGYFYLAPKVDGTLCSPDSTSVQVQCKIRKAGCDNGLGSKRRFKCGVCGGDNKSK 683
DB 649 GIGYFVYLQPKVDGTLPCSPDSTSVQVQCVKRGACDRIIDSKKPKDKCGVCGGNGSTCK 708
QY 684 KVTGLTFKPHGYNFVVAIPAGASSIDIRORYGKGLIGDDNYIALKNQSKYLLNGHFVV 743
DB 709 KISGSVTSAPGYHDIITPTGATNIEVKORNGSRNNGSFLAIRAADGTIILNGDYTL 768
QY 744 SAVERDLVWKSLLRYSSTGTAVESLQASRPILFVTLVSVGKMTPPRVYSFYLKPE 803

DB 769 STLEQDIMYKGVLYRSGSSAALERINSPLEPTIQVLTVGNALRPAIKYTYFVKK 828
QY 804 PREDKSSHPKDPGRPSVLIHNSVLSLSNQVEQDPPARVAGSWGPCSAACSGSLQKRA 863
DB 829 -----KES-----FNAITFS-----AWVIEWGECSKSCBLGWORRL 861
QY 864 VDCRGSAGORTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSCGRGFORSLSKC 920
DB 862 VECRDINGQ---PASECAKEVPKASTRPPCADHPCEPQWOLGSEWSSCKTCGKGYKRSKLC 918
QY 921 VGHGRLARDQCNLHRKPOE-LDFCVLRPC 950
DB 919 LSHDGGVLSHESCDPLKPKHFDICTMAEC 949
RESULT 8
AAB50002
ID AAB50002 standard; Protein; 950 AA.
XX
AC AAB50002;
XX
DT 19-MAR-2001 (first entry)
XX
DE Human METH1.
XX
KW Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiofibroma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control.
XX
OS Homo sapiens.
XX
PN WO2000071577-A1.
XX
PD 30-NOV-2000.
XX
PF 25-MAY-2000; 2000WO-US14462.
XX
PR 25-MAY-1999; 99US-0318208.
PR 20-JUL-1999; 99US-0144882.
PR 10-AUG-1999; 99US-0147823.
PR 13-AUG-1999; 99US-0373658.
PR 22-DEC-1999; 99US-0171503.
PR 22-FEB-2000; 2000US-0183792.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (SMK) SMITHKLINE BEECHAM CORP.
PA (BETH) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
PA (JONA/) JONAK Z L.
PA (TRUL/) TRULLI S H.
PA (FORN/) FORNWALD J A.
PA (TERR/) TERRETT J A.
XX
PI IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Fornwald JA, Terrett JA;
XX
XX WPI; 2001-025136/03.
DR N-PSDB; AAC90057.
XX
PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX

Claim 15; Fig 1; 768pp; English.

PS The present sequence is human METH1 (ME for metalloprotease and TH for thrombospondin). METH1 can be used for inhibiting angiogenesis in an individual, and for treating cancer, benign tumours, an ocular angiogenic disease, rheumatoid arthritis, psoriasis, delayed wound healing, nonunion fractures, scleroderma, trachoma, vascular adhesions, myocardial angiogenesis, coronary collaterals, cerebral collaterals, arteriovenous malformations, ischaemic limb angiogenesis, Osler-Weber syndrome, plaque neovascularisation, telangiectasia, haemophilic joints, angiofibroma, fibromuscular dysplasia, wound granulation, Crohn's disease or atherosclerosis. METH1 can also be used in birth control. METH1 can also be used in diagnostic methods for the prognosis of cancer.

SQ. Sequence 950 AA;

Query Match 48.0%; Score 2480.5; DB 22; Length 950;

Best Local Similarity 48.9%; Pred. No. 6.8e-184; Mismatches 251; Indels 101; Gaps 24;

Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
 DB LLLLAALLAVSDALGRPSDEDELVP-ELE-----RAP---GHGTRRLRLHAF 64

QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSDLRRCFTYSGDVNAEPDSF 114
 DB DQDLDELRLPDSFLAPGFTLQNGRKSGETPLP-----ETDLAHCFYSGTVNGDPSSA 119

QY 115 AAVSLCGLGAGFYGAEVISPLNAS---APAAQRNSQGA---HLLQ---RRGVPG 165
 DB AALSCEGVGAFYLLGEAYFIQPLPAASERLATAAPGKPPAPLOFHLRRNRQGDVG 179

QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGEGSRSSRS 207
 DB TCGVVDDEPRPTGKAETDEDEGTEGDEGPQMS-----FQDFALQGVGP-TGTGS 230

QY 208 GRAKRFVSIPIRYVELTVVADSMVKFHGADLEHYLTLLATAARLYRHPSILNPINIVV 267
 DB IRKRRFVSHRYVETMLVADQSMAEFHGSLKHYLLTLFSAARLYKHPISIRNSILVVV 290

QY 268 KVLLLRDSDGPKVTGNAALTLNFCNAOKLKNVSKDKHPEYWDTAILFTRQDLGATTC 327
 DB KILVIHQEGPEVTSNAALTLNFCNWKQHNPPSDRDAEHYDTALFTRQDLGSGTQC 350

QY 328 DTLMADVGMCDPKRSCVIEDDGLPSAFTTAHELGHVFNMPHDNVKCEEVFGKLRAN 387
 DB DTLMADVGMCDPKRSCVIEDDGLQAAFTTAHELGHVFNMPHDNAKQACASLNGVQDS 410

QY 388 HMSPTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPISLPDLPGASVYLSQQ 447
 DB HMASMLSLNDHSQWPSPCSAYMITSLDNGHGBCLMDKPNPQLPGDLPGTSYDANRQ 470

QY 448 CELAFGVGSRPCP-YMOYCTKLACTGAKGMVQCTRHFPWADGTSCGEGKILKGACVE 506
 DB COFTGEDSKHCPDAASTCTSLWCTGTSGVGLVQCTRHFPWADGTSCGEGKINGKCVN 530

QY 507 RHNLNKH---RVDGSAKWDPYPCSTCGGVQOLARQCTNPTPANGKYCEGVRYKYR 563
 DB KTD-RKHFDTFHSGWGWGWDGSCSTCGGVQVYTHRECDNVPKNGKYCEGVRYR 589

QY 564 SCNLEPCSSASGSKFREOCFAFNPNHSTNRLTLAVAVPKYSGVSPKDKCKLICRAN 623
 DB SCNLEDCPDN-NGKTFREOCEAHNEFSKASFGSGPAAVEIPKYGVSPPKDKCKLICQAK 648

QY 624 GTGYFYVLAPVVDGTLCSDPSTSVCGVKCIKAGCDNGLSKKRFDKCGVCGDNKSK 683
 DB GIGYFFVLQPKVVDGTFPCSPDSTSVCGQCVKAGCDRIIDSKKKFDKCGVCGGNGSTCK 708

QY 684 KVTGLFKPMHYNFVVAIPAGASSIDIRORYKGLIGDDNYLALKNSQKYLINGHFVV 743
 DB KISGSVTSKPGYHDIITPGATNIEVKQRNQRSGSRNNGSFLAKAADGTIILNGDYTL 768

RESULT 9
 AAW80285
 ID AAW80285 standard; Protein; 967 AA.

XX AAW80285;
 AC 19-JAN-1999 (first entry)
 DT Human integrin ligand polypeptide ITGL-TSP.
 XX ITGL-TSP; integrin ligand; angiogenic disease; cancer; atherosclerosis;
 KW chronic inflammatory disorder; rheumatoid arthritis; tissue remodeling;
 KW macular degeneration; diabetic retinopathy; Alzheimer's disease; human;
 KW restenosis.
 OS Homo sapiens.
 XX EP874050-A2.
 XX 28-OCT-1998.
 PD 27-JAN-1998; 98EP-0300575.
 XX 24-APR-1997; 97US-0845496.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX Fronwald JA, Hastings GA, Jonak ZL, Terrett JA;
 PI Trulli SH;
 XX WPI; 1998-54643/47.
 DR N-PSDB; AAW66508.
 XX DNA encoding integrin ligand polypeptide ITGL-TSP - used to treat
 PT angiogenic diseases, restenosis, Alzheimer's disease and in tissue
 PT remodeling
 XX Claim 11; Pages 6-9; 24pp; English.
 PS This represents a human integrin ligand polypeptide ITGL-TSP. ITGL-TSP
 CC polypeptides can be used in the treatment of angiogenic diseases such as
 CC cancer, cancer metastasis, chronic inflammatory disorders, rheumatoid
 CC arthritis, atherosclerosis, macular degeneration or diabetic retinopathy,
 CC restenosis, Alzheimer's disease and tissue remodeling. They can be used
 CC to treat a subject in need of enhanced activity or expression of the
 CC ITGL-TSP polypeptide.
 XX Sequence 967 AA;
 SQ Query Match 48.0%; Score 2480.5; DB 19; Length 967;
 Best Local Similarity 48.9%; Pred. No. 7e-184;
 Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAF 60

Db 36 LLLAAALVSDALGRSEDEELVWP-ELE-----RAP-----GHCTTRLRHAF 81
QY 61 QEDFVHLHTPDQAFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCPYSGDVNAEPDSF 114
Db 82 DQDLLELRPSSFLAPFTLQNGVRKSGSETPLP-----ETDLAHCFYSGTVNGDPSSA 136
QY 115 AAVSLCGGLRGAFYRGAYEYVISPUNAS-----APAAORNSOGA-----HLQO--RRGVPGG 165
Db 137 AALSCEGVGAFFYLLGAYFIQPLPAASERLATAAPGKPPALQFHLRRNQGVGG 196
QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGFSGSRRRS 207
Db 197 TCGVVDDEPRPTGKAETDEDEGTBEGDEGPQWS-----PQDPAQGVGPQ-TGTGS 247
QY 208 GRAKEFVSIPRYVETLVVADSMVKFHGADLEHYLLTLLATAARLYRHPSTILNINIVV 267
Db 248 IRKKFVSSHRYVETLVVADSMVAFHGSLGHKLHLLTFSVAARLYRHPSTIRNSVSLVV 307
QY 268 KYLLLRDRDSGPKVTGNAALTLRNFCAWQKKLVSKVSKHPEYWDTAILFTRODLGATTC 327
Db 308 KILVTHDEQKPEVTSNAALTLRNFCAWQKHNPSPDRDAEHYDTAILFTRODLGSGQTC 367
QY 328 DTGLNADVTWCDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDNNVKKVCEVEFGKLRAN 387
Db 368 DTGLNADVTWCDPKRSCSVIEDGLQAAFTTAHELGHVFNPHDNNVKKVCEVEFGKLRAN 427
QY 388 HMSPTLLQIDRANPWSACSAITITDLDGSHGDCLLDOPSKPISLPELPGASYTILSQO 447
Db 428 HMASLSNLDHSPWSPCSAYMTSFLDNGHGECIMDKPQNPQLPGLDLPSTIDANRQ 487
QY 448 CELAFGVSKPCP-YMQVYTKLWCTGKAKQWCVQTRHFPWADGTCGEGKCLKLGACVE 506
Db 488 COFTFGEDSKPCPAAASTCSTLWCTGSGVLVQCTKHFHWADGTCGEGKWCINGKCVN 547
QY 507 RHNLKH---RVDGSWAKWDPYKSCSTCGGVQOLARQCTNPTTPANGKYCEGVRYR 563
Db 548 KTD-RKHEDTFFHSGWGMWGPWCGDCSTCGGVQYTWRECDNPVPKNGKYCEGVRYR 606
QY 564 SCNLEPCPSSASGKSFREEQEAFNGYNHSTNRLTLAVAWVPKYSGVSPRDKCLICRAN 623
Db 607 SCNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSGPAVEWIPKYAGVSPDKRCKLICQAK 665
QY 624 GTGYFVVLAPKVVDGTLCPDSTSVCGQKIKAGCDGNLGSKRFPKCGVCGGDNKSK 683
Db 666 GIGYFFVLQPKVVDGTLCPDSTSVCGQKIKAGCDRIIDSKKFKDKCGVCGGNGSTCK 725
QY 684 KVTGLFKPMHGVNFVAIPAGASSIDIRORYKGLIGDNDYLAALKNSQKYLINGHFVV 743
Db 726 KISGVSFAKPGYHDIITPTGATNIEVKQNRGSRNNGSFSLAIAKAADGTYLINGDYTL 785
QY 744 SAVERDLVVGSLRLYSGTGTAVESLQASRPILFLETVLSVGMKPTPRVRYSFYLPKE 803
Db 786 STLEQDIMYKGVLRYSGSSAALIRSFSPLEPLTIQVLTGVALRPAIKYTFVKK 845
QY 804 PREDKSSHPKDPGCVSLHNSVLSLQVQPDPRPARVAGSWGSPCSAGSGLOKRA 863
Db 846 ----KES-----FNAIPTFS-----AWVIEWEGCSKSCSLGQORRL 878
QY 864 VDCRSAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGFORSLKC 920
Db 879 VECRDINGQ---PASECAKEVKPASTRPCADHPCPQWLGENSSCSTCKGKYKRSLSKC 935
QY 921 VGHGRLLRDQCNLHRRKQOE-LDFCVLRPC 950
Db 936 LSHDGGVLSHESCDPLKKPKHFIDFCTMAEC 966

RESULT 10
AAB50011
ID AAB50011 standard; Protein: 968 AA.
XX
AC AAB50011;

XX
DT 19-MAR-2001 (first entry)
XX DE Protein; SEQ ID 125.
XX DE
KW Human; METH1: metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiofibroma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control.
XX
OS Homo sapiens.
XX
PN WO200071577-A1.
XX
PD 30-NOV-2000.
XX
PF 25-MAY-2000; 2000WO-US14462.
XX
PR 25-MAY-1999; 99US-0318208.
PR 20-JUL-1999; 99US-0144882.
PR 10-AUG-1999; 99US-0147823.
PR 13-AUG-1999; 99US-0373658.
PR 22-DEC-1999; 99US-0171503.
PR 22-FEB-2000; 2000US-0183792.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (SMK) SMITHKLINE BEECHAM CORP.
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
PA (JONA/) JONAK Z L.
PA (TRUL/) TRULLI S H.
PA (FORN/) FORNWALD J A.
PA (TERR/) TERRETT J A.
XX
PI Truela-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Fornwald JA, Terrett JA;
XX
DR WPI; 2001-025136/03.
XX
PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX
PS Claim 15; Pages 759-763; 768pp; English.
XX
CC The present invention relates to human METH1 and METH2 (ME for
CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC METH can be used for inhibiting angiogenesis in an individual, and for
CC treating cancer, benign tumours, an ocular angiogenic disease,
CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH can also be used in birth control. METH can also
CC be used in diagnostic methods for the prognosis of cancer. The present
CC sequence is a protein isolated in the present invention.
XX
SQ Sequence 968 AA;
Query Match 48.0%; Score 2480.5; DB 22; Length 968;
Best Local Similarity 48.9%; Pred. No. 7e-184;
Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY 1 MLLIGILTLAFAAGTAGGPEREVVYVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 DB 37 LLLLAALLAVSALGRPSDEELVVP-ELE-----RAP---GHGTRRLRHAF 82
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRCFYSGDVNAPDSF 114
 DB 83 DQQLDLRLPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFTSGTVNGDPSSA 137
 QY 115 AAVSLCGGLRGAGYRGAEVVISPLPNAS---APAAQRNSOGA-----HLLQ---RRGVPGG 165
 DB 138 AALSCEGVRYGAYFIQPLPAASERLATAAPCEKPPAPLOFHLRLRRNRQGDVG 197
 QY 166 PSG-----DPTSRC-----GVASG---WNPAILRALDPYKPRRAGFGESRRRS 207
 DB 198 TCGVVDDEPRPTGKAETDEDETEGEDEGPQWS-----PQDPALQGVGP-TGTGS 248
 QY 208 GRAKREVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPISILNINIVV 267
 DB 249 IRKRFVSSHRYVETMLVADQSMAEFHGSLGHYLLTLFSVAARLYKHPISIRNSVLVV 308
 QY 268 KVLRLDRDSDGPKVTGNAALTNRNFCAMOKKLNKVSCKHPEYWDTAILFTRODLGCATTC 327
 DB 309 KILVIHQEGPEVTSNAALTNRNFCAMOKHNPSPDRDAEHYDTAILFTRODLGSGQTC 368
 QY 328 DTLCGMADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVENMPHDNVKVEEYFGKLRAN 387
 DB 369 DTLCGMADVGVCDPSRSCSVIEDDGLQAFTTAHELGHVENMPHDNAKOCASLNGYNQDS 428
 QY 388 HMSPTLIQIDRANPWSACSAAITDPLDSHGDCLLDQPSKPISELPEDLPASYSYLSQ 447
 DB 429 HMASMLNLDHSPQSPCSAYMITSPLDNGHCECLMDKPNQPLQPLDPLGTSYDANRQ 488
 QY 448 CELAFGVGSKPCP-YMOCYTKLACTGKAKOMVCONPHEPMDAGTSCGEGKCLKACAYE 506
 DB 489 CQFTFGSDSKHCPDAACTSLMCTGTSGVGVLCQTHFPAWDTSCGEGKWCINGKCVN 548
 QY 507 RHNLNKH---RVDSWAKWDPYPCSTRCGGVQGLARROCTNTPANGGKYCEGVRYKYR 563
 DB 549 KTD-RKHFDTPFHSGWGMWPGWDCSRTCGGVQVYTMRECDNPVPKNGGKYCEGVRYR 607
 QY 564 SCNLEPCSSASGSKFEEQCEAFNGYHNSTNRLTLAVAVPKYSGVSPDKCKLCIRAN 623
 DB 608 SCNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSGPAVEVPIKYGVPKDRCKLICQAK 666
 QY 624 GTGYFYVLAKPVVDGTLCSPDSTSVQGVQKICIRAGCDNGLGSKRRFDKCGVCGDNKSKC 683
 DB 667 GIGYFFVLPKPVVDGTPCSPDSTSVQGVQCVKAGCDRIIDSKKKFDKCGVCGNGSTCK 726
 QY 684 KVTGLTKPMHGYNVVAIPAGASSIDIRORYKGLTGDDNYIALKNSQGYLLNGHFVV 743
 DB 727 KISGVSATKPGYHDIITPGATNIEVKORNGRNRNGSFLAIAKAADGYIILNGDYTL 786
 QY 744 SAVERDLVWGLSLRYSGTAVESLOASRPILEPLVEVLSVGMKTPPRVRYSFYLPKE 803
 DB 787 STLEQDLMKGVVLRYSGSSAALRIRSFPLKEPLTIQVLTGNALRPKIKITFYVKKK 846
 QY 804 PREDKSHPKDRGPSVLHNSVLSLSNQVEQDPRPARWAGSWGSPCSASCSSGSLQKRA 863
 DB 847 ---KES-----FNAIPTFS-----ANVIEWGCSSKSELGWQRRL 879
 QY 864 VDCRGSAQRTVPACDAAH---RPVETQACGE-PCPTWELSAWSPCSKSGRGQRRLSKC 920
 DB 880 VECRDINGQ---PASECAKEVKPAPTRPCADHPQCPQWOLGEWSSCKTKCGKYKRSKLC 936
 QY 921 VGHGRLIARDQCNLHRRKPOE-LDFCVLRLPC 950
 DB 937 LSHDGGVLSHESCDPLKPKHFIDFCTMAEC 967

RESULT 11
 AAY04142
 ID AAY04142 standard; Protein; 967 AA.

XX AAY04142;
 AC 15-JUN-1999 (first entry)
 DT Human Tango-71 protein.
 DE Human; Tango-71; Tango-73; Tango-74; Tango-76; Tango-83; diagnosis;
 KW detection.
 KW Homo sapiens.
 OS WO9907850-AL.
 PN 18-FEB-1999.
 PD 06-AUG-1998; 98WO-US16502.
 PF 05-SEP-1997; 97US-0058108.
 PR 06-AUG-1997; 97US-0054966.
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 XX Goodearl ADJ, Holtzman DA;
 PI WPI; 1999-167426/14.
 DR N-PSDB; AAX19955.
 DR XX
 PT New TANGO polypeptides and nucleic acids encoding them - useful as
 PT diagnostic agents and for treating disorders caused by aberrant
 PT expression of TANGO
 XX Claim 8; Fig 1; 84pp; English.
 PS The present sequence represents human Tango-71. Tango polypeptides are
 CC useful for identifying compounds which bind the polypeptide via direct
 CC binding, competition binding assays or Tango-71, -73, -74, 76 or -83-
 CC mediated signal transduction. Tango polypeptides are also useful for
 CC identifying modulating compounds by determining effect on Tango activity.
 CC Tango polypeptides and nucleic acids are useful for diagnosing diseases
 CC related to aberrant expression of Tango, and Tango polypeptides are
 CC useful for raising antibodies which can be used in diagnostic assays for
 CC detection of Tango, and also for generating anti-idiotypic antibodies for
 CC prevention and protection.
 XX SQ Sequence 967 AA;
 Query Match 48.0%; Score 2479.5; DB 20; Length 967;
 Best Local Similarity 48.9%; Pred. No. 8.4e-184;
 Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;
 QY 1 MLLIGILTLAFAAGTAGGPEREVVYVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 DB 36 LLLLAALLAVSALGRPSDEELVVP-ELE-----RAP---GHGTRRLRHAF 81
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRCFYSGDVNAPDSF 114
 DB 82 DQQLDLRLPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFTSGTVNGDPSSA 136
 QY 115 AAVSLCGGLRGAGYRGAEVVISPLPNAS---APAAQRNSOGA-----HLLQ---RRGVPGG 165
 DB 137 AALSCEGVRYGAYFIQPLPAASERLATAAPCEKPPAPLOFHLRLRRNRQGDVG 196
 QY 166 PSG-----DPTSRC-----GVASG---WNPAILRALDPYKPRRAGFGESRRRS 207
 DB 197 TCGVVDDEPRPTGKAETDEDETEGEDEGPQWS-----PQDPALQGVGP-TGTGS 247
 QY 208 GRAKREVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPISILNINIVV 267
 DB 248 IRKRFVSSHRYVETMLVADQSMAEFHGSLGHYLLTLFSVAARLYKHPISIRNSVLVV 307
 QY 268 KVLRLDRDSDGPKVTGNAALTNRNFCAMOKKLNKVSCKHPEYWDTAILFTRODLGCATTC 327

Db 308 KILVIHQEGPEVTSNAALTRNFCNMQKQHNPPSDRDAEHYDAILTRQDLGSGQTC 367
 QY 328 DTLMADVGTMCDPKRSRCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRAN 387
 Db 368 DTLMADVGTCDPSCRSVIEDDGLQAFATTAHELGHVFNPHDDAKQACASLNGVNDSD 427
 QY 388 HIMSPTLIOIDRANPWSACSAIITFDLSDGHGDCLLDOPSKPISLPEDLPASVYTLSSQ 447
 Db 428 HMASMLNLDHSDQSPSPCSAYMITSFLDNGHGECLMDKPQNPIQLPGLDLPSTSDANRQ 487
 QY 448 CELAFGVGSKPCP-YMQVCTKLWCTKAKGOMWCOTRFPFPADGTCGEGKLCAGACVE 506
 Db 488 COFTGEDSKHCPDAASTCTLMWCTSGGVLVCOCTKFPWADGTCGEGKNGKCVN 547
 QY 507 RHNLANH---RVDSWAKWDPTGPCSRCTCGGVQLARQCCTNPTPANGKYCEGVRYK 563
 Db 548 KTD-RKHFDTPPHSGMWGPGWDCSRCTCGGVQYTMRECDNPVPKNGKYCEGKRVYR 606
 QY 564 SCNLEPCPSSASGKSFREOCEAFNGYHNHSTNRLTLAVAWVPKYSGVSPDKCKLICRAN 623
 Db 607 SCNLEDCPDN-NGKTFREOCEAHNEFSKASFGSPAVIEWIPKYAGVSPDKCKLICQAK 665
 QY 624 GTGYFVLAPKVDGTLCSPDSTSVCGQKICAGCDNGLSKKREDKCGVCGDNKSKC 683
 Db 666 GIGYFVLQPKVDGTLCSPDSTSVCGQCVKAGCDRIIDSKKRFDKCGVCGGNGSTCK 725
 QY 684 KVTGLTFPMHGVNFWAIPAGASSIDIRQRYKGLIGDDNYLAKNSQKRYLLNGHFVV 743
 Db 726 KISGVTSAKPYHDIITPGATNIEVKQNRGSRNNGSLAIAKADGTYILNGDYTL 785
 QY 744 SAVERDLVVKGLRYSGTGTAVESLQASRIEPLIVEVLVSGKMPPRVRSFYLPKE 803
 Db 786 STLEDIMYKGVLUKYSGAALIRSFSLKEPLTQVLTGNALRPKIKTYFYVKK 845
 QY 804 PREDKSHPKDPRGSPVLHNSVLNSQVEQDPRPARWAGSWGPCSCGSGLOKRA 863
 Db 846 ---KES-----FNAIPFS-----AWIEWEGCSKSELGWQRLL 878
 QY 864 VDCRGSAGQRTVPACDAH--RPVETQACGE-PCPTWELSAWSPKSCGGRFORSLKC 920
 Db 879 VECRDINGQ---PASECAKEYKPASTRPCADHPQCPQWLGEWSSCSCTCGKYKRSKLC 935
 QY 921 VGHGRLILDQCNLHRKPQE-LDFCVLRPC 950
 Db 936 LSHDGGVLSHESCDPLKKPKHFIDFCTMAEC 966

RESULT 12
 AAW78189
 XX AAW78189 standard; Protein; 967 AA.
 AC AAW78189;
 XX
 DT 13-APR-1999 (first entry)
 XX Human secreted protein encoded by gene 64 clone HOUQ17.
 DE
 XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 40
 FT /label= unknown
 FT Misc-difference 45
 FT /label= unknown

FT Misc-difference 169 /label= unknown
 FT Misc-difference 293 /label= unknown
 FT Misc-difference 297 /label= unknown
 FT Misc-difference 557 /label= unknown
 FT /label= unknown
 XX
 PN W09856804-A1.
 XX
 XX 17-DEC-1998.
 XX
 PF 11-JUN-1998; 98WO-US12125.
 XX 02-OCT-1997; 97US-0061060.
 PR 13-JUN-1997; 97US-0049547.
 PR 13-JUN-1997; 97US-0049548.
 PR 13-JUN-1997; 97US-0049549.
 PR 13-JUN-1997; 97US-0049550.
 PR 13-JUN-1997; 97US-0049606.
 PR 13-JUN-1997; 97US-0049607.
 PR 13-JUN-1997; 97US-0049608.
 PR 13-JUN-1997; 97US-0049609.
 PR 13-JUN-1997; 97US-0049610.
 PR 13-JUN-1997; 97US-0049611.
 PR 13-JUN-1997; 97US-0050566.
 PR 13-JUN-1997; 97US-0050901.
 PR 13-JUN-1997; 97US-0052989.
 PR 08-JUL-1997; 97US-0051919.
 PR 18-AUG-1997; 97US-0055984.
 PR 12-SEP-1997; 97US-0058665.
 PR 12-SEP-1997; 97US-0058668.
 PR 12-SEP-1997; 97US-0058669.
 PR 12-SEP-1997; 97US-0058750.
 PR 12-SEP-1997; 97US-0058971.
 PR 12-SEP-1997; 97US-0058972.
 PR 12-SEP-1997; 97US-0058975.
 PR 02-OCT-1997; 97US-0060834.
 PR 02-OCT-1997; 97US-0060841.
 PR 02-OCT-1997; 97US-0060844.
 PR 02-OCT-1997; 97US-0060865.
 PR 02-OCT-1997; 97US-0061059.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Ebner R, Ferrie AM, Feng P, Greene JM, Lafleur DW;
 PI Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P;
 PI Yu GL;
 XX
 DR WPI: 1999-080881/07.
 DR N-PSDB; AAX04374.
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 11; Page 297-300; 380pp; English.
 XX
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AAX04302) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 86 novel genes and their fragments (nucleic
 CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 86
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX04311 for described uses).

XX	Sequence	967 AA;	
SO	Query Match	47.6%; Score 2458.5; DB 20; Length 967;	
	Best Local Similarity	48.5%; Pred. No. 3.6e-182;	
	Matches	481; Conservative 154; Mismatches 255; Indels 101; Gaps 24;	
QY	1	MLLGILTLAFAGTAGGFFPEREVVPIRLDPDINGRRYYWRGPDSDGQGLIFQITAF	60
DB	36	LLLLAALLXVSDALGRSDEDELVVP-ELE-----RAP---GHGTRLRHAF	81
QY	61	QEDFYLHLTPDAQFLAPAFSTEHGLG-----VPLQGLTGGSSDLRRCFYSGDVNAEPDSF	114
DB	82	DQDLDELRLPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFYSTVNGDPSSA	136
QY	115	AAVSLCGGLRAGFYGAEGYVISPDPNAS---APAAQRNSOGA-----HLIQ--RRGVPGG	165
DB	137	AALSCEGVGARGFYLLGEAYFIQPLPAASERLXTAAPGERKPPAPLQPHLLRRNRQGDVG	196
QY	166	PSG-----DPTSRC-----GVASG--WNPAILRALDPVKPRRAGFGESSRRRS	207
DB	197	TCGVVDDEPRPTGRAETEDDEGTGEGDEGPQWS-----FQDPALQGVGP-TGTGS	247
QY	208	GRAKRFVSIPIRYVETLVVADESVMKFFGADLEHYLLTLATAARLYRHPHSILNPINIVVY	267
DB	248	IRKRFVSSHRYVETMLVADQSMAEFHGSLKHYLLTLFSAARLXKHPXIRNSVSLVVY	307
QY	268	KVLLLRDRSGPKVTGNAAITLRFNFCAWOKLNVSKDHPYWDATLFTRODLGGATTC	327
DB	308	KILVIHQKGPVETSNAAITLRFNFCWQKQHNPPSDORDAEHYDTALFTRODLGGATTC	367
QY	328	DTLGMADVTGMDCKRSCVTEDDGLPSAFTTAHELGHVFNMPHDNVKCEEVFEGKLRAN	387
DB	368	DTLGMADVTGMDCKRSCVTEDDGLPSAFTTAHELGHVFNMPHDNVKCEEVFEGKLRAN	427
QY	388	HMSPTLIQIDRANPWSACSAITITFDLSHGBCLLDQPSKPSLSPEDLPASITLSQQ	447
DB	428	HMASMLNLDHSPWSPSCSAIMTISFLDNHGECCLMDKQNPQLPGLPGLTSDYANRQ	487
QY	448	CELAFGVGSKPCP-YMQYCTKTLWCTGAKGQWVQTRHPWADGTSGEGKCLCLKACVE	506
DB	498	COFTGEDSKPCDPAASTCTSLWCTGTSGVLVQCTRHPWADGTSGEGKWCINGKCVX	547
QY	507	RHNLNKH---RVDSNAKWDYGPSCSRTCGGVQVLARQCTNPANGKCYCEGVRYKYR	563
DB	548	KTD-RKHEDTFFHSGWMGWPWGDGCSRTCGGVQVYTRCEDNPPVKNKGKCYCEGVRYR	606
QY	564	SNLEPCSSASGKSFREOCEAFNGYHNSTNRLTLAVAWPKYSGVSPRDKCLICRAN	623
DB	607	SNLEPCSSASGKSFREOCEAFNGYHNSTNRLTLAVAWPKYSGVSPRDKCLICRAN	665
QY	624	GTGYFVVLAPKVDGTLGSPDSTSVYQVQKCIKAGCDNLSKSRFDCKVCYCGGDNKSK	683
DB	666	GIGYFVVLAPKVDGTLGSPDSTSVYQVQKCIKAGCDNLSKSRFDCKVCYCGGDNKSK	725
QY	684	KVTGLFTPMHGYNTVVAIPAGASSIDTRQYKGLICDDNVLKNSQKYLINGHVVV	743
DB	726	KISGVSITAKPGYHDIITPTGATNIEVKQNRQNRSGFLAIAKAADGTIYILNGDYL	785
QY	744	SAVERDLVVKGLLRYSGTGAVESLQASRPITLPLEVLSVSGKMTPPRVRYSYFLPKPE	803
DB	786	STLEQDQWYKGVVLYSGSSAALEIRFSPLKEPLTLOVLTGVNALRPKIYFYFKK	845
QY	804	PREDKSSHHPKPRGPSVLHNSVLSLNSQVQPDPRPPARWAGSPGCSASCGLQKRA	863
DB	846	RES-----FNAIPFS-----AWVEENGESKSCGLGWQRRL	878
QY	864	VDCRSAGORTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCKSCGRGFORRSKUC	920
DB	879	VECRDINGQ---PASECAKEVAPSTRPCADHPQWQWLGWSSCKTCGKGYKRSUKC	935
QY	921	VGHGGRLLARDQCNLHRKPQE-LDFCVLRPC	950

DB	936	LSHDGGVLSHESCDPLKPKKHFDFTMAEC	966
RESULT 13			
AA53899			
ID	AA53899	standard; Protein: 950 AA.	
XX	AC	AA53899;	
DT	13-MAR-2000	(first entry)	
XX	DE	Amino acid sequence of a murine ADAMTS-1 protein.	
XX	XX	GON-1; metalloproteinase; cell migration; modulator; ADAMTS-1;	
KW	KW	metalloproteinase domain; thrombospondin domain; abnormal cell migration;	
KW	KW	organ shaping; sterility; cancer metastasis.	
XX	OS	Mus sp.	
PN	PN	WO9961656-A2.	
XX	PD	02-DEC-1999.	
XX	PF	28-MAY-1999; 99WO-US11918.	
XX	PR	29-MAY-1998; 98US-0087170.	
PR	PR	13-APR-1999; 99US-0129023.	
XX	XX	(WISC) WISCONSIN ALUMNI RES FOUND.	
PA	PA	Kimble JE, Bielloch RH;	
PI	PI	WPI; 2000-072633/06.	
XX	DR	Identifying modulators of proteins containing metalloprotease and	
XX	PT	thrombospondin domains, potentially useful for controlling cell	
PT	PT	migration and organ shaping	
XX	XX	Disclosure; Fig 1C; 60pp; English.	
PS	XX	The present sequence represents a murine ADAMTS-1 protein. ADAMTS-1 is	
CC	CC	a metalloproteinase. The specification describes another related	
CC	CC	metalloproteinase, a C. elegans GON-1 protein, that lacks a transmembrane	
CC	CC	domain and possesses a predicted metalloproteinase domain between residues	
CC	CC	269-456. In C. elegans hermaphrodites, GON-1 is required for migration of	
CC	CC	two distal tip cells to produce elongated tubes, whereas in males, GON-1	
CC	CC	is required for migration of a single linker cell to produce a single	
CC	CC	elongated tube. The protein is used in the method of the invention. The	
CC	CC	specification describes a method for identifying a modulator of a	
CC	CC	protein that contains a metalloproteinase domain and a thrombospondin	
CC	CC	domain. The method comprises treating a target organism, having a	
CC	CC	developing gonadal cell that is responsive to the protein, with a test	
CC	CC	compound, and determining any change in migration or shape of the cell	
CC	CC	attributable to the test compound. The compounds identified are	
CC	CC	potential therapeutic modulators of abnormal cell migration and organ	
CC	CC	shaping, e.g. for rendering animals (specifically nematodes) sterile	
CC	CC	and for inhibiting cancer metastases.	
XX	XX	Sequence 950 AA;	
QY	1	MLLIGILTLAFAGTAGG--FEPERVVVPIRLDPDINGRRYYWRGP-EDSGDGLIFOI	57
DB	20	LLLASITMLLCARGAHGRETEDELVLP-SLE-----RAPGHDSITTRL--RL	66
QY	58	TAQEDFYLHLTPDAQFLAPAFSTEHGLVPLQGLTGGSS-----SDLRRCFYSGDVNA	109
DB	67	DAFQQQLHLKLPDPSGLAPGFTLQTV-----GRSPGSEAHQHLDPGDLAHCFYSTVNG	121
QY	110	EPDSFAVAVSLGGRLGARGFYGAEGYVISPDPNAS-----APAAQRNSOGA-----HLQRRG	161

```
Db 122 DPGSAAALSLCEGVGAFYLOGEFFIOPAPGVATERLAPAVPEESSARPQFHILRR- 180
QY 162 VFGGSGDPTSCGVASGWNPAILRALDPYKPRRAGFESRR- 206
Db 181 ----RGSGGAKGVMD-----DETLPDSRPESQNTNRNQPVRDPTPOACKP 226
QY 207 ----SGRAKRFVSPRYVETLVVADESVMVKFHGADLEHYLLTLATAARLYRHPSTLNPI 262
Db 227 SGPGRSTRKRFVSSPRYVETMLVADQSMADFHGSLGKHYLLTLFSAARFYKHPSTRNSI 286
QY 263 NIWVKVLLLRDSDGPKVTGNAALTNRNFCANOKLKNKYSKDKHPEYWDTAILFTRODLC 322
Db 287 SLVVVKILYIEQKGPVETSNAAALTNRNFCANOKLKNKYSKDKHPEYWDTAILFTRODLC 346
QY 323 GATTCDTLGMADVGTMCDPKRSVCVIEDDGLPSAFTTAHELGHVFNPHNDKVKVCBEVFG 382
Db 347 GSHTCDTLGMADVGTMCDPKRSVCVIEDDGLPSAFTTAHELGHVFNPHNDKVKVCBEVFG 405
QY 383 KLRANHMSPITLIOIDRANPWSACSAALITDFLDSHGDCLLDOPSKPISLEPDLPGASY 442
Db 406 VSGDHLKASMLSSLDHSPQSPCAVMTVTFDNGHGECIMDKPQNPRLPSDLPGLTY 465
QY 443 TLSQCELAFAVGSKPCP-YMQYCTKLWCTGKAGQMVCTRFPWADGTSCEGKLCILK 501
Db 466 DANROCCFTFGESKHCPCDAASTCTLWCTGSLGVLVCCCTKHPWADGTSCEGKWCVS 525
QY 502 GACVERHNLKH---RVDGSAKWDYPGCSRTCGGVQOLARRQCTNPTPANGKYCEGV 558
Db 526 GKCVMKTD-KHFAFVHSGVPMGPMGDCSRTCGGVQVMTRECDNPPVKMGKYCEGK 584
QY 559 RVKRSCLNLEPCSSAGSKSPREQCAFNGYHSTNRLTLAVAWPKYSGVSPRDKCL 618
Db 585 RVYRSCNIEDCPDN-NGKTFRESCRAHNEFSKASFGNEPTVETWPKYAGVSPRDKCL 643
QY 619 ICRANGTYFYVLAKVVDGTLCPDSTSVQVQKCIKACDGNLGSKKRFPKCGVCGGD 678
Db 644 TCEAKGIGYFFVLOPKVVDGTPCSPDSTSVQVQKCIKACDRIIDSKKKFCKGVCYCGN 703
QY 679 NKSCKVTKGLFTKPMHGVVVAIPAGASSIDIRQGYKGLIGDNDYALKNSQGRYLLN 738
Db 704 GSTCKKMSGIYSTPRGYHDIVTIPAGATNIEVKHNRQGRNRNGSFLAIRADGTIILN 763
QY 739 GHFVVSVERDLVKGSLRLRSVGTGAVESLQASRPILPTEVLSVCKMTPPVRYSF 798
Db 764 GNFTLSTLEQDLYKGTLYRSGSAALEIRSFSLPELTLTQVLMVGHALPKIKFTY 823
QY 799 YLPKPEPREDKSHPKDPRGSPVLSNLSNQVEQPDPRPARWAGVSGWPCSCSGSG 858
Db 824 FMKKKTES-----FNAIPTES-----EWIEEWGECSTCGSG 856
QY 859 LQKRAVDCRGSAQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGFOR 915
Db 857 WQRRVVQCRDINGH---PASECAKEVPKASTRCPADLPCHPHQVGMVQWSPCSKTCGKGYK 913
QY 916 R-SLKCVGHGGRLLARDQCNLHKKPQE-LDFCVLRPC 950
Db 914 RCTLKCVDHGGVLSNCSCLPKPKPKHYIDFCLITQC 950
```

RESULT 14

```
AAB21265
ID AAB21265 standard; Protein; 896 AA.
XX
XX AAB21265;
XX
XX
DT 23-FEB-2001 (first entry)
XX
XX Mouse metalloproteinase ADAMTS-1.
XX
XX Mouse; ADAMTS-1; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; nontropic; neuroprotective; antiparkinsonian;
```

```
KW cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
OS Mus musculus.
XX
XX W0200053774-A2.
XX
XX 14-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US06237.
XX
XX 08-MAR-1999; 99US-0264585.
XX
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
XX Kelner GS, Clark M, Maki RA;
XX
XX WPI; 2000-594326/56.
XX
XX Polynucleotide encoding novel members of a disintegrin,
XX metalloproteinase and thrombospondin domain protein family used to
XX prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Disclosure; Fig 17; 129pp; English.
XX
XX The present sequence is mouse metalloproteinase ADAMTS-1. The
XX ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
XX and Metalloproteinase domain) family. Members of the ADAMTS family
XX contain a thrombospondin domain in addition to the disintegrin and
XX metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
XX useful for the manufacture of medicaments for treating conditions
XX associated with neuroinflammation and/or neurodegeneration, such as
XX Alzheimer's disease, Parkinson's disease and stroke. They are also
XX useful for treating conditions associated with cell proliferation, cell
XX migration, inflammation and/or angiogenesis, such as cancer, arthritis
XX and autoimmune diseases. They can be used to treat patients afflicted
XX with an invasive tumour, a brain tumour or brain injury.
```

Query Match 896 AA;

```
Best Local Similarity 45.0%; Score 2322; DB 21; Length 896;
Matches 453; Conservative 143; Mismatches 236; Indels 110; Gaps 21;
QY 1 MLLGILTLAFAGTAGG--FEPEREVYVPIRLDPDINGRYYWRGP-EDSGDQGLIFI 57
Db 20 LLLASITMLLCARGAHGRPTDEELVLP-SLE-----RAPGHDSTTTL--RL 66
QY 58 TAFQEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGS-----SDLRRCFYSGDVNA 109
Db 67 DAFGQQLHLKLPDQSGFLAPGFTLV-----GRSPGSEAOHLDPDGLAHCFYSGTVNG 121
QY 110 EPDSFAAVSLCGGLRGAFGYRGAEYVISPFPNAS-----APAAQRNSOGA----HLQRRG 161
Db 122 DPGSAAALSLCEGVGAFYLOGEFFIOPAPGVATERLAPAVPEESSARPQFHILRR- 180
QY 162 VFGGSGDPTSCGVASGWNPAILRALDPYKPRRAGFESRR- 206
Db 181 ----RGSGGAKGVMD-----DETLPDSRPESQNTNRNQPVRDPTPOACKP 226
QY 207 ----SGRAKRFVSPRYVETLVVADESVMVKFHGADLEHYLLTLATAARLYRHPSTLNPI 262
Db 227 SGPGRSTRKRFVSSPRYVETMLVADQSMADFHGSLGKHYLLTLFSAARFYKHPSTRNSI 286
QY 263 NIWVKVLLLRDSDGPKVTGNAALTNRNFCANOKLKNKYSKDKHPEYWDTAILFTRODLC 322
Db 287 SLVVVKILYIEQKGPVETSNAAALTNRNFCANOKLKNKYSKDKHPEYWDTAILFTRODLC 346
QY 323 GATTCDTLGMADVGTMCDPKRSVCVIEDDGLPSAFTTAHELGHVFNPHNDKVKVCBEVFG 382
Db 347 GSHTCDTLGMADVGTMCDPKRSVCVIEDDGLPSAFTTAHELGHVFNPHNDKVKVCBEVFG 406
```

```

QY 383 KLRANHMSTLTQIDRANPWSACSAIITDFLDSHGDCILDDQPSKPISLPDLPGASY 442
Db 407 VTGDSHLMASLWSLSDHSPWSPSCSAIYMTVTFDNGHGECCLMDKQNPITKLPDLPLGLY 466
QY 443 TLSQOCLAFGYSKPCP-YMOYCTKLWCTGAKGOMVOTRHFWDGTCGEGKCLK 501
Db 467 DANRQOFTGEGSKPCDAASTCTTLWCTGTSGLLVCOVTHFHWADGTCGEGKWCVS 526
QY 502 GACVERHNLNKH--RVDSWAKWDPYGPCSRTCGGGVOLARRQCTNTPPANGKYCBGV 558
Db 527 GKCVNKTDM-KHEATPVHSGWPGWPGWDCSRTCGGGVQYTMRECDNVPKNGKRCBGK 585
QY 559 RVYRSCNLPSPSSASGSEFEQCEAFNGYNHSTNRLTLAVAKVPKYSVSPEDCKL 618
Db 586 RVIRSCNIEDCPDN-NGKTFREEQCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKL 644
QY 619 ICRANGTYEYVLAPKVPDGTGLCSPDSTSVCGQKICAKGCDNGLGSKKRFDKCGVCGD 678
Db 645 TCEAKGIGYFVLQPKVWDGTGCPSPDSTSVCGQCVKAGCDRIIDSKKKFDKCGVCGN 704
QY 679 NKCKKVTGLFTKPMHGVFVVAIPAGASSIDIRORYKGLIGDDNYLALNKSOGKYLIN 738
Db 705 GSTCKHSGVITSTRGYHDIVTPAGATNIEVKHNRQNGSRNNGSFLAIRAADGTIIN 764
QY 739 GHFVSAVERDLVKGSLRYSGTGPVSLQASRPILPLVEVLSVGMKTPPRVRYSF 798
Db 765 GNETSLTLEQDLYKGTVLRYSGSSAALERIRSFPLKEPLTIQVLMVGHALRPKFTY 824
QY 799 YLKEPREDKSSHPKDPGRPSVLHNSVLISNQVEQDPRPAPVAGSWGSPCSASCSSG 858
Db 825 FMKKKTES-----FNAIPTFS-----EWIEWEGCSKTCSSG 857
QY 859 LQKRAVDCRSAGQRTVPACDAAH--RPVETQACGE-PCPTW 897
Db 858 WQRRVVOQRDINGH--PASECAKEVKPASTRCPADLPDPCHW 896

RESULT 15
AAW78435
ID AAW78435 standard; Protein; 727 AA.
XX
AC AAW78435:
XX
DT 11-MAY-1999 (first entry)
XX
DE Human ADAMTS-1 protein.
XX
KW Metalloproteinase-disintegrin protein with a thrombospondin domain; ARDS;
KW ADAMTS-1; drug composition; foodstuff; leukocyte; thrombocyte; hepatitis;
KW blood; erythrocyte; inflammatory disease; rheumatoid arthritis; asthma;
KW nephritis; Crohn's disease; acute respiratory disease syndrome.
XX
OS Homo sapiens.
XX
PN W09855643-A1.
XX
PD 10-DEC-1998.
XX
PF 03-JUN-1998; 98WO-JP02449.
XX
PR 03-JUN-1997; 97JP-0160422.
XX
PA (KURE ) KUREHA CHEM IND CO LTD.
XX
PI Hakozaki M, Hirose K, Inoguchi E, Ishida Y, Ishioka K;
PI Kuno K, Matsushima K;
XX
DR WPI; 1999-070277/06.
DR N-PSDB; AAX17990.
XX
PT Human metalloproteinase-disintegrin protein with thrombospondin
PT domain - useful as leukocyte and thrombocyte decreasing and
PT erythrocyte increasing agent

```

```

XX
PS
CC This sequence represents a novel human metalloproteinase-disintegrin
CC protein with a thrombospondin domain (ADAMTS-1). The protein may be used
CC in drug compositions and foodstuffs, as an agent for decreasing the
CC leukocyte and thrombocyte blood count and increasing the erythrocyte
CC blood count, e.g. for treatment of inflammatory diseases such as
CC rheumatoid arthritis, hepatitis, nephritis, Crohn's disease, asthma
CC and ARDS.
XX
SQ Sequence 727 AA;
Query Match 44.0%; Score 2274; DB 20; Length 727;
Best Local Similarity 54.4%; Pred. No. 5.4e-168; Indels 40; Gaps 11;
Matches 408; Conservative 129; Mismatches 173;
QY 209 RAKFEFYSIPRYETLVVADESVMVKFHGADLEHYLLTLAATARLYRHPSILNPINVVVK 268
Db 9 RKREFVSSPRYETLVVAADQSNAEFHSGLKHLLTFSVAARLYKHPISIRNSVSLVVVK 68
QY 269 VLLLRDSDGPKVTGNAALTILNFCWAKKLNKVSQKHPEYWDTAILETRQDLGATCD 328
Db 69 ILVIHDEQKGPEVTSNAALTILNFCWQKHNPSSDRDAEHYDTAILETRQDLGSGQTC 128
QY 329 TLGNADVGTMCDPKRCSVIEDDGLPSAFTTAHELGHVFNPHDNVVKCEEVFGKLRANH 388
Db 129 TLGNADVGTVCDPSRCSVIEDDGLQAFTTAHELGHVFNPHDDAKQASINGVNQDSH 188
QY 389 MMSPTLIQIDRANPWSACSAIITDFLDSHGDCILDDQPSKPISLPDLPGASYTLSSOC 448
Db 189 MMASMLNSLDHSPWSPSCSAIYMTVTFDNGHGECCLMDKQNPITKLPDLPLGLYDANRQC 248
QY 449 ELAFGVGSKPCP-YMOYCTKLWCTGAKGOMVOTRHFHWADGTCGEGKCLKAGACVER 507
Db 249 QFTFGEDSKHCPDAASTCTLWCTGTSGLVLVQCTKHFPWADGTCGEGKCKNGKCVNK 308
QY 508 HNLNKH--RVDSWAKWDPYGPCSRTCGGGVOLARRQCTNTPPANGKYCBGVKRYRS 564
Db 309 TD-RKHFDTPFHSGWPGWPGWDCSRTCGGGVQYTMRECDNVPKNGKYCBGKRVIRTS 367
QY 565 CNLEPCSSASGSKSFEQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRCKKLCIRANG 624
Db 368 CNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSPGPAVEWIPKYAGVSPKCKLTCQAKG 426
QY 625 TGYEYVLAPKVPDGTGLCSPDSTSVCGQKICAKGCDNGLGSKKRFDKCGVCGDNKSKK 684
Db 427 IGYEYVLQPKVWDGTGCPSPDSTSVCGQCVKAGCDRIIDSKKKFDKCGVCGNGSGTCKK 486
QY 685 VTGLFTKPMHGVFVVAIPAGASSIDIRORYKGLIGDDNYLALNKSOGKYLINAGHFVYS 744
Db 487 ISGSVTSAPGYHDIVTIPTGATNIEVKORNGSRNNGSFLAIRAADGTIILNGDYTL 546
QY 745 AVERDLVVGSLRYSGTGPVSLQASRPILPLVEVLSVGMKTPPRVRYSFYLPREP 804
Db 547 TLEQDINRYKGVLLRYSGSSAALERIRSFPLKEPLTIQVLMVGNALRPKIKYTFVKKK- 605
QY 805 REDKSSHPKDPGRPSVLHNSVLISNQVEQDPRPAPVAGSWGSPCSASCSSGLQKRAV 864
Db 606 ---KES-----FNAIPTFS-----AWIEWEGCSKSCELGWQRRLV 639
QY 865 DCRGSAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRQFQRLSKCV 921
Db 640 ECRDINGQ---PASECAKEVKPASTRCPADHPQWQLGEMSSCKTCTGKGYKKRSLKCL 696
QY 922 GHGRLRLARDQCNLHRKQPE-LDFCVLRPC 950
Db 697 SHDGGVLSHESCDPLKKPKHFIDFCTLTQC 726

```

Search completed: April 29, 2003, 17:20:27
Job time : 61.2301 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:14:33 ; Search time 56 seconds
(without alignments)
3495.442 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 950
Sequence: 1 MLLGLITLAFAGTAGGFE.....DQCNLHRKPOELDFCVLRPC 950

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_virus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	931	98.0	950	4 Q8TE58	Q8TE58 homo sapien
2	71	7.5	340	11 Q91256	Q91256 mus musculus
3	10	1.1	192	6 Q95N24	Q95N24 equus caball
4	10	1.1	1072	4 Q8TE57	Q8TE57 homo sapien
5	10	1.1	1159	4 Q8TEY8	Q8TEY8 homo sapien
6	10	1.1	1223	4 Q8WXS8	Q8WXS8 homo sapien
7	10	1.1	1223	4 Q8TE55	Q8TE55 homo sapien
8	10	1.1	1235	4 Q95428	Q95428 homo sapien
9	10	1.1	1572	5 Q44938	Q44938 haemonchus
10	10	1.1	2165	5 Q19791	Q19791 caenorhabdi
11	10	1.1	3198	5 Q9U868	Q9U868 manduca sex
12	9	0.9	269	6 Q9GL54	Q9GL54 oryctolagus
13	9	0.9	1081	4 Q8TE60	Q8TE60 homo sapien
14	9	0.9	1280	11 Q9EPX2	Q9EPX2 mus musculus
15	9	0.9	2174	5 Q9GQR0	Q9GQR0 drosophila
16	9	0.9	3060	5 Q9VAV4	Q9VAV4 drosophila

```

17      8      0.8      101      8      099713
18      8      0.8      105      6      09GLK7
19      8      0.8      117      16      08XND1
20      8      0.8      120      16      09KYP8
21      8      0.8      122      16      092M53
22      8      0.8      192      11      09CX59
23      8      0.8      196      16      08Z6C7
24      8      0.8      196      16      034160
25      8      0.8      197      17      08TSC3
26      8      0.8      206      5      024927
27      8      0.8      381      8      033604
28      8      0.8      407      13      092032
29      8      0.8      408      10      094LF5
30      8      0.8      409      13      08OG89
31      8      0.8      422      5      09YIV4
32      8      0.8      428      5      09YIV5
33      8      0.8      434      16      099XT2
34      8      0.8      440      16      097SS4
35      8      0.8      456      2      093D60
36      8      0.8      466      5      017451
37      8      0.8      482      13      09PVK9
38      8      0.8      525      4      096RW4
39      8      0.8      539      5      08WPC7
40      8      0.8      605      8      094WQ9
41      8      0.8      607      8      09XL01
42      8      0.8      608      16      09PMJ4
43      8      0.8      687      4      09BXR8
44      8      0.8      712      5      043981
45      8      0.8      713      10      09MAC6

```

ALIGNMENTS

```

RESULT 1
Q8TE58      PRELIMINARY;      PRT;      950 AA.
ID          Q8TE58;
AC          Q8TE58;
DT          01-JUN-2002 (TrEMBLrel. 21, Created)
DT          01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT          01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE          Metalloprotease disintegrin 15 with thrombospondin domains.
GN          ADAMTS15.
OS          Homo sapiens (Human).
OC          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX          NCBI_TaxID=9606;
RN          [1]
RP          SEQUENCE FROM N.A.
RX          MEDLINE=21856482; PubMed=11867212;
RA          Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA          Lopez-Otin C.;
RT          "Cloning, expression analysis, and structural characterization of
RT          seven novel human ADAMTS, a family of metalloproteinases with
RT          disintegrin and thrombospondin-1 domains.";
RL          Gene 283:49-62(2002).
DR          EMBL; AJ315733; CAC86014.1;
KW          Integrin; Protease.
SQ          SEQUENCE: 950 AA; 103286 MW; 5DFBE18285CCCC3B CRC64;

Query Match      98.0%; Score 931; DB 4; Length 950;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 931; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      20 EPREVVVPIRLDPDINGRIYWRGPDSDGQGLIFQITAFQDFYVHLTPDQAFAPAF 79
      |||||||
Db      20 EPREVVVPIRLDPDINGRIYWRGPDSDGQGLIFQITAFQDFYVHLTPDQAFAPAF 79
      |||||||
QY      80 STEHLGVPLQGLTGGSDLLRRCFYSGDVNAEPDSFAVSLCGGLRCAGFYRGAEYISPL 139
      |||||||
Db      80 STEHLGVPLQGLTGGSDLLRRCFYSGDVNAEPDSFAVSLCGGLRCAGFYRGAEYISPL 139
      |||||||
QY      140 PNASAPAAQRNSQGAHLLQRRGVPGGPGSDPTSRCGVASGWNPAILRALDYPKPRRAGFG 199

```

```
|||||
Db 140 PNASAPAAQRNSQAHLLQRRGYPGGSGDPTSRGCVASGWNFAILLRALDYPKPRAGFC 199
Qy 200 ESRSSRRSGRAKRVSTIPRVETLVVADESMVKFPGADLEHYLLTLTAAARLYRHPSIL 259
Db 200 ESRSSRRSGRAKRVSTIPRVETLVVADESMVKFPGADLEHYLLTLTAAARLYRHPSIL 259
Qy 260 NPINIVVVKVLLLRDRSGPKVGTGNAALTLRNFCANQKLNKYSDKHPEYWDTAILLTRQ 319
Db 260 NPINIVVVKVLLLRDRSGPKVGTGNAALTLRNFCANQKLNKYSDKHPEYWDTAILLTRQ 319
Qy 320 DLGATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVCEE 379
Db 320 DLGATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVCEE 379
Qy 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIITDFLDSHGDCCLLDQPSKPISLPDLPG 439
Db 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIITDFLDSHGDCCLLDQPSKPISLPDLPG 439
Qy 440 ASYTLSSQCELAFGVSGKPCPYMOYCTKLWCTGKAGQMVQCTRHFPWADGTSCGEGKLC 499
Db 440 ASYTLSSQCELAFGVSGKPCPYMOYCTKLWCTGKAGQMVQCTRHFPWADGTSCGEGKLC 499
Qy 500 LKGACVERHNLNKHVRDGSMAKMDPYGPCSRTCGGGVQLARRQCTNPTPANGGKYCEGVR 559
Db 500 LKGACVERHNLNKHVRDGSMAKMDPYGPCSRTCGGGVQLARRQCTNPTPANGGKYCEGVR 559
Qy 560 VKYRSCNLEPCPSSASGKSFRBEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDCKLI 619
Db 560 VKYRSCNLEPCPSSASGKSFRBEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDCKLI 619
Qy 620 CRANGTGYFYVLAPEVVDGTLCSPDSTSVCVQKCIKAGCDGNLGSKKRDKCGVCGGDN 679
Db 620 CRANGTGYFYVLAPEVVDGTLCSPDSTSVCVQKCIKAGCDGNLGSKKRDKCGVCGGDN 679
Qy 680 KSKKKVTGLFTKPMHGVNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNSSQGYLLNG 739
Db 680 KSKKKVTGLFTKPMHGVNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNSSQGYLLNG 739
Qy 740 HFVYSAVERDLVKGSLRLRSGTGTAVESLQASRPILPTVLSVGVKMTPPRVRYSFY 799
Db 740 HFVYSAVERDLVKGSLRLRSGTGTAVESLQASRPILPTVLSVGVKMTPPRVRYSFY 799
Qy 800 LPKEPREDKSHPKDRGPSVYLNHNSVLSLNOVEQPDPRPPARVAGSWGPCSCSGSL 859
Db 800 LPKEPREDKSHPKDRGPSVYLNHNSVLSLNOVEQPDPRPPARVAGSWGPCSCSGSL 859
Qy 860 QKRAVDCRGSAQRTVPACDAHRPVETQACGEPCTWELSAWSPCKSCGCGFRRSLK 919
Db 860 QKRAVDCRGSAQRTVPACDAHRPVETQACGEPCTWELSAWSPCKSCGCGFRRSLK 919
Qy 920 CVHGGRLRLARDQCNLHRKPOELDFCVLRPC 950
Db 920 CVHGGRLRLARDQCNLHRKPOELDFCVLRPC 950

RESULT 2
Q91Z56 PRELIMINARY; PRT; 340 AA.
ID Q91Z56;
AC Q91Z56;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DR 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE Similar to a disintegrin and metalloproteinase with thrombospondin
motifs 1 (ADAMTS-1) (Fragment).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

Query Match 1.1%; Score 71; DB 11; Length 340;
Best Local Similarity 100.0%; Pred. No. 1.1e-67; Indels 0; Gaps 0;
Matches 71; Conservative 0; Mismatches 0;

Qy 685 VTGLETKPMHGVNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNSSQGYLLNGHFVYS 744
Db 75 VTGLETKPMHGVNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNSSQGYLLNGHFVYS 134

Qy 745 AVERDLVWVKS 755
Db 135 AVERDLVWVKS 145

RESULT 3
Q95N24 PRELIMINARY; PRT; 192 AA.
ID Q95N24;
AC Q95N24;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DR 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE Aggreacanase-1 (Fragment).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Flannery C.R., Little C.B.;
RT "Expression and activity of equine aggreacanases.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF368321; AAK53425.1; -
DR MEROPS; M12.221; -
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000130; Zn.MTpeptidse.
DR PROSITE; PS50215; ADAM_MEPPO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 192
SQ SEQUENCE 192 AA; 20670 MW; 9013B0E19FCE8C56 CRC64;

Query Match 1.1%; Score 10; DB 6; Length 192;
Best Local Similarity 100.0%; Pred. No. 0.091; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

Qy 360 AHELGHVFN 369
Db 29 AHELGHVFN 38

RESULT 4
Q8TE57 PRELIMINARY; PRT; 1072 AA.
ID Q8TE57;
AC Q8TE57;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DR 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE Metalloprotease disintegrin 16 with thrombospondin type I motif.
GN ADAMTS16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
```



```

RA Madan A., Dickhoff R., Shaffer T., James R., Lasky S., Hood L.;
RT "Complete sequence of the gene for presenilin 1.";
RL EMBL; AF109907; AAC97963.1; -.
DR HSP; P12111; 2KNT.
DR InterPro; IPR003598; Iq_c2.
DR InterPro; IPR003006; Iq_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00047; Iq; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSPI; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 2.
DR PROSITE; PS00092; TSPI; 4.
KW Hypothetical protein; Immunoglobulin domain;
KW Serine protease inhibitor.
SQ SEQUENCE 1235 AA; 133477 MW; A0B44CCE4F38E350 CRC64;

Query Match 1.1%; Score 10; DB 4; Length 1235;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 527 PCSRTCGGV 536
DB 37 PCSRTCGGV 46
|||||

RESULT 9
O44938 PRELIMINARY; PRT; 1572 AA.
AC O44938;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Thrombospondin.
GN THRI.
OS Haemonchus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOREDUN;
RA Skuce P.J., Newlands G.F.J., Stewart M., Pettit D., Smith D.,
RT Knox D.P.;
RT "Cloning and characterization of thrombospondin, a novel multidomain
RT glycoprotein associated with the gut of Haemonchus contortus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043121; AAB99830.2; -.
DR HSP; P05067; ICAO.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00014; Kunitz_BPTI; 6.
DR Pfam; PF00090; tsp_1; 6.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 6.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00131; KU; 6.
DR SMART; SM00209; TSPI; 7.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 6.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 6.
DR PROSITE; PS00092; TSPI; 2.
KW Serine protease inhibitor.
SQ SEQUENCE 1572 AA; 171871 MW; 2260B30DC2F903EC CRC64;

Query Match 1.1%; Score 10; DB 5; Length 1572;

```

```

Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGVQ 537
DB 83 CSRTCGGVQ 92
|||||

RESULT 10
Q19791 PRELIMINARY; PRT; 2165 AA.
ID Q19791; Q27524;
AC Q19791; Q27524;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Cajadast S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smaison M., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1; -.
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1; -.
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSP; P15167; 1DTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTpeptidse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 14.
DR SMART; SM00209; TSPI; 18.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPI; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;

Query Match 1.1%; Score 10; DB 5; Length 2165;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGVQ 537
DB 615 CSRTCGGVQ 624
|||||

RESULT 11
Q908G8 PRELIMINARY; PRT; 3198 AA.
ID Q908G8
AC Q908G8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

```

01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE lacunin precursor.
OS Manduca sexta (Tobacco hawkmoth) (Tracheata; Hexapoda; Insecta);
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Spingioidea; Spingidae; Spinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99457716; PubMed=10528409;
RA Nardi J.B., Martos R., Walder K.K., Lampe D.J., Robertson H.M.;
RT "Expression of lacunin, a large multidomain extracellular matrix
protein, accompanies morphogenesis of epithelial monolayers in Manduca
sexta.";
RL Insect Biochem. Mol. Biol. 29:883-897(1999).
DR EMBL; AF078161; AAF04457.1; -.
DR HSP; P12111; 2KNT.
DR InterPro; IPR004094; Antistatin.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR002221; WAP.
DR Pfam; PF02822; Antistatin; 4.
DR Pfam; PF00047; Ig; 2.
DR Pfam; PF00014; Kunitz_BPTI; 10.
DR Pfam; PF00095; WAP; 1.
DR PRINTS; PR00759; BASICPTASE.
DR PRODOM; PD000222; Kunitz_BPTI; 10.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00131; KU; 10.
DR SMART; SM00209; TSPl; 7.
DR SMART; SM00217; WAP; 1.
DR PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 8.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 10.
DR PROSITE; PS00092; TSPl; 1.
KW Immunoglobulin domain; Serine protease inhibitor; Signal.
FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 3198 AA; 349364 MW; AB4ACD459C0D9134 CRC64;
Query Match 1.1%; Score 10; DB 5; Length 3198;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 527 PCSRTCGGV 536
DB 71 PCSRTCGGV 80
ID Q9GL54 PRELIMINARY; PRT; 269 AA.
AC Q9GL54;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Aggrecanase-2 (Fragment).
GN ADAMTS-11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX Goad D.L., Goad M.E.;
RT "Molecular cloning of lapine aggrecanase-2 (ADAMTS-11) from articular
chondrocytes.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF317415; AAG33062.1; -.
DR HSP; O9PW35; 1BUD.
DR MEROPS; M12.225; -.
DR InterPro; IPR001590; Reprolysin.

DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00209; TSPl; 1.
DR PROSITE; PS02115; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPl; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1 269
FT NON_TER 269 269
SQ SEQUENCE 269 AA; 29193 MW; 97A1CA80B33452FA CRC64;
Query Match 0.9%; Score 9; DB 6; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 329 TLGMADVGT 337
DB 1 TLGMADVGT 9
ID Q8TE60 PRELIMINARY; PRT; 1081 AA.
AC Q8TE60;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ADAMTS18 protein.
GN ADAMTS18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
seven novel human ADAMTSs, a family of metalloproteinases with
disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ311903; CAC83612.1; -.
SQ SEQUENCE 1081 AA; 119655 MW; 0438BF645676461E CRC64;
Query Match 0.9%; Score 9; DB 4; Length 1081;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 528 CSRTCGGV 536
DB 600 CSRTCGGV 608
ID Q9EPX2 PRELIMINARY; PRT; 1280 AA.
AC Q9EPX2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Papilin.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CD-1;
RL MEDLINE=20530499; PubMed=11076767;
RA Krameroval A.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Krameroval A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,

RA Steron A.L., Prockop D.J., Fessler J.H.;
 RT "Papilin in development; a pericellular protein with a homology to the
 RL ADAMTS metalloproteinases";
 RL Development 127:5475-5485(2000).
 DR EMBL: AF314171; AAG41980.1; -
 DR HSP: P12111; 2KNT
 DR InterPro: IPR000183; Decarboxylase2.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003600; Ig_like.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR000884; TSPI.
 DR Pfam: PF00047; Ig; 3.
 DR Pfam: PF00014; Kunitz_BPTI; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR PRODOM: PD000222; Kunitz_BPTI; 1.
 DR SMART: SM00409; IG; 3.
 DR SMART: SM00408; Igc2; 3.
 DR SMART: SM00410; IG_like; 2.
 DR SMART: SM00131; KU; 1.
 DR SMART: SM00209; TSPI; 5.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 1.
 DR PROSITE: PS00879; ODR_DC_2_2; UNKNOWN_1.
 DR PROSITE: PS00092; TSPI; 3.
 KW Immunoglobulin domain; Serine protease inhibitor.
 SQ SEQUENCE 1280 AA; 138824 MW; AE287705E561AF30 CRC64;

Query Match 0.9%; Score 9; DB 11; Length 1280;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 527 PCSRTCGGG 535
 |||||
 Db 38 PCSRTCGGG 46

RESULT 15

Q9GORO PRELIMINARY; PRT; 2174 AA.
 AC Q9GORO;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Extracellular matrix protein papilin precursor.
 GN PPN OR CG1540 OR CG18436.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DP CN BW;
 RX MEDLINE=20530499; PubMed=11076767;
 RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
 RA Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,
 RA Steron A.L., Prockop D.J., Fessler J.H.;
 RT "Papilin in development; a pericellular protein with a homology to the
 RL ADAMTS metalloproteinases";
 RL Development 127:5475-5485(2000).
 DR EMBL: AF205357; AAG37995.1; -
 DR HSP: P12111; 2KNT
 DR FlyBase: FBgn003137; Ppn.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003598; Ig_c2.
 DR InterPro: IPR003600; Ig_like.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR000884; TSPI.

DR InterPro: IPR002221; WAP.
 DR Pfam: PF00047; Ig; 2.
 DR Pfam: PF00014; Kunitz_BPTI; 3.
 DR Pfam: PF00090; tsp.1; 5.
 DR Pfam: PF00095; wap; 1.
 DR PRINTS: PR00003; 4DISULPHCORE.
 DR PRODOM: PD000222; Kunitz_BPTI; 3.
 DR SMART: SM00409; IG; 3.
 DR SMART: SM00408; Igc2; 2.
 DR SMART: SM00410; IG_like; 1.
 DR SMART: SM00131; KU; 3.
 DR SMART: SM00209; TSPI; 7.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS00317; 4DISULFIDE_CORE; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 3.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 3.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS00092; TSPI; 3.
 KW Matrix protein; Serine protease inhibitor; Signal.
 FT SIGNAL 1 26
 SQ SEQUENCE 2174 AA; 231936 MW; 038F707952623120 CRC64;

Query Match 0.9%; Score 9; DB 5; Length 2174;
 Best Local Similarity 100.0%; Pred. No. 9.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGGV 536
 |||||
 Db 69 CSRTCGGGV 77

Search completed: April 29, 2003, 17:23:43
 Job time: 67 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:22:09 ; Search time 27 seconds
(without alignments)

3382.507 Million cell updates/sec

Title: US-10-009-332-1

Perfect score: 950

Sequence: 1 MLLLGILTAFAAGTAGGFE.....DQCNLHRKPQELDFCVLRPC 950

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR.73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	1.8	951	2 T00017	gene ADAMTS-1 prot
2	14	1.5	550	2 T47158	hypothetical prote
3	13	1.4	837	2 T00355	hypothetical prote
4	10	1.1	1205	2 T18517	procollagen N-endo
5	10	1.1	2165	2 T21371	hypothetical prote
6	8	0.8	196	2 AF0715	probable ABC trans
7	8	0.8	206	2 A45517	coccidioidis-relate
8	8	0.8	407	2 S66260	metalloproteinase
9	8	0.8	440	2 D95029	PTS system, IIC co
10	8	0.8	440	2 G97900	hypothetical prote
11	8	0.8	607	2 T11032	NADH2 dehydrogenas
12	8	0.8	608	2 A81293	probable flagellar
13	8	0.8	712	2 A45638	immunodominant mic
14	8	0.8	788	2 T25061	hypothetical prote
15	8	0.8	1049	2 T30525	alpha-mannosidase
16	8	0.8	5037	1 A54161	ryanodine-binding
17	7	0.7	66	2 AB2371	hypothetical prote
18	7	0.7	96	2 I68623	neurofibromin - mo
19	7	0.7	108	2 A12333	hypothetical prote
20	7	0.7	110	2 C72472	hypothetical prote
21	7	0.7	111	2 B26567	nitrogen regulator
22	7	0.7	112	2 S13078	nitrogen regulator
23	7	0.7	112	2 S52328	nitrogen regulator
24	7	0.7	112	2 C87493	nitrogen regulator
25	7	0.7	112	2 D82985	nitrogen regulator
26	7	0.7	112	2 AB2794	nitrogen regulator
27	7	0.7	112	2 AD3374	nitrogen regulator
28	7	0.7	119	2 G81738	hypothetical prote
29	7	0.7	124	2 AH5226	conserved hypothet

30	7	0.7	125	2 A83048	hypothetical prote
31	7	0.7	131	1 D69827	hypothetical prote
32	7	0.7	132	2 A10244	probable rhodanese
33	7	0.7	135	1 B49205	virulence-associat
34	7	0.7	138	2 B70867	hypothetical prote
35	7	0.7	142	2 C70059	hypothetical prote
36	7	0.7	143	2 E64359	ribosomal protein
37	7	0.7	147	2 S51810	trypsin inhibitor
38	7	0.7	157	2 A97573	nitrogen regulator
39	7	0.7	160	2 G71122	hypothetical prote
40	7	0.7	161	2 T11855	DnaJ protein homol
41	7	0.7	163	2 B70559	hypothetical prote
42	7	0.7	167	2 E87476	conserved hypothet
43	7	0.7	172	2 JH0780	trypsin inhibitor
44	7	0.7	172	2 A24082	trypsin inhibitor
45	7	0.7	172	2 A27220	trypsin inhibitor

ALIGNMENTS

RESULT 1

T00017
gene ADAMTS-1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00017
R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.
Genomics 46, 466-471, 1997
A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1
A:Reference number: Z14055; MUID:98110583; PMID:9441751
A:Accession: T00017
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-951 <KUN>
A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BA24501.1; PID:g2809057
A:Experimental source: strain 129SVJ
C:Genetics:
A:Gene: ADAMTS-1
A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 603/1; 660/3; 719/2
C:Superfamily: thrombospondin type 1 repeat homology
F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 1.8%; Score 17; DB 2; Length 951;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNPHD 372

|||||

Db 380 AFTTAHELGHVFNPHD 396

RESULT 2

T47158
hypothetical protein DKFZp762C1110.1 - human (fragment)

C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47158
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weill, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24379
A:Accession: T47158
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-550 <AAA>

A:Cross-references: EMBL:AL162080

A:Experimental source: adult melanoma (MeWo cell line); clone DKFZp762C1110
C:Genetics:
A:Note: DKFZp762C1110.1

Query Match 1.5%; Score 14; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 484 HFPWADTSGEGK 497
|
Db 108 HFPWADTSGEGK 121

RESULT 3

T00355
hypothetical protein KIAA0688 - human
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00355
R: Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5: 169-176, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: Z14142; MUID:98403880; PMID:9734811
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-837 <ISH>
A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BAA31663.1; PID:g3327190
A:Experimental source: Brain
C:Genetics:
A:Gene: KIAA0688
C:Superfamily: thrombospondin type 1 repeat homology
F:515-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 1.48; Score 13; DB 2; Length 837;
Best Local Similarity 100.0%; Pred. No. 0.00023; Indels 0; Gaps 0;
Matches 13; Conservative 0; Mismatches 0;

QY 311 DTAILFTRQDLGG 323
|
Db 311 DTAILFTRQDLGG 323

RESULT 4

T18517
collagen N-endopeptidase (EC 3.4.24.14) I - bovine
N:Alternate names: procollagen N-proteinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18517
R: Colige, A.; Nusgens, B.V.; Lapiere, C.M.
submitted to the EMBL Data Library, February 1996
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A:Reference number: Z18941
A:Accession: T18517
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-1205 <COL>
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1
A:Experimental source: skin
C:Genetics:
A:Gene: PC I-NP
C:Function:
A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior to
C:Keywords: hydrolase; metalloproteinase

Query Match 1.1%; Score 10; DB 2; Length 1205;
Best Local Similarity 100.0%; Pred. No. 0.39; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

QY 670 DRKGVCGGDN 679
|
Db 704 DRKGVCGGDN 713

RESULT 5

T21371
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T21371; T24896
R: Gajadsty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone F25H8

R: Gajadsty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
C:Genetics:
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1;

Query Match 1.1%; Score 10; DB 2; Length 2165;
Best Local Similarity 100.0%; Pred. No. 0.63; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

QY 528 CSRTCGG3VQ 537
|
Db 615 CSRTCGG3VQ 624

RESULT 6

AF0715
probable ABC transport ATP-binding chain STY1861 [imported] - Salmonella enterica sub
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AF0715
R: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Conington, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
S.; Moulie, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quall, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AF0715
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-196 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD02095.1; PID:g16502930; GSPDB:GN00176
C:Genetics:
A:Gene: STY1861

Query Match 0.8%; Score 8; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 9.8; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

QY 183 AILRALDP 190
|
Db 131 AILRALDP 138

RESULT 7

A45517
cocciidiosis-related antigen - Eimeria tenella (fragment)
N:Alternate names: thrombospondin-related antigen, 100K
C:Species: Eimeria tenella
C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 07-May-1999
C:Accession: A45517; S27818
R: Clarke, L.E.; Tomley, F.M.; Wisner, M.H.; Foulds, I.J.; Boursnell, M.E.G.
Mol. Biochem. Parasitol. 41, 269-280, 1990

A:Title: Regions of an Eimeria tenella antigen contain sequences which are conserved in
A:Reference number: A45517; MUID:90377296; PMID:2204833
A:Accession: A45517
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <CIA>
A:Cross-references: GB:M32988; NID:gl58870; PID:gl58871
A:Superfamily: thrombospondin type 1 repeat homology
F:136-198/Domain: thrombospondin type 1 repeat homology <THR5>

Query Match 0.8%; Score 8; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTC GGG 535
|||||
DB 88 CSRTC GGG 95

RESULT 8
S66260
A:Title: Molecular cloning and sequence analysis of cDNAs for metalloproteinases from br
A:Reference number: S66259; MUID:95314311; PMID:7793974
A:Accession: S66260
A:Molecule type: mRNA
A:Residues: 1-407 <SEL>
A:Cross-references: EMBL:U18234; NID:g603216; PIDN:AA59704.1; PID:g603217
A:Accession: S74263
A:Molecule type: protein
A:Residues: 188-206 <SEA>
A:Superfamily: atrolysin C
A:Keywords: hydrolase; metalloproteinase; zinc
F:118/Domain: signal sequence #status predicted <SIG>
F:19-187/Domain: propeptide #status predicted <PRO>
F:188-407/Product: metalloproteinase #status experimental <MAT>
F:304-384,344-351,346-368/Disulfide bonds: #status predicted
F:329,333,339/Binding site: zinc, catalytic (His) #status predicted
F:330/Active site: Glu #status predicted

Query Match 0.8%; Score 8; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 VGTMC DPK 342
|||||
DB 300 VGTMC DPK 307

RESULT 9
D95029
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: D95029
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74429.1; PID:gl4971719; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4

PTB system, IIC component [imported] - Streptococcus pneumoniae (strain TIGR4)
A:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
A:Accession: D95029
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: D95029
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74429.1; PID:gl4971719; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4

C:Genetics:
A:Gene: SP0250
C:Superfamily: phosphotransferase system enzyme II factor II, phosphoenolpyruvate-de

Query Match 0.8%; Score 8; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 VPLOGLTG 93
|||||
DB 226 VPLOGLTG 233

RESULT 10
G97900
A:Title: hypothetical protein PTS-EIIC [imported] - Streptococcus pneumoniae (strain R6)
A:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
A:Accession: G97900
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G97900
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99035.1; PID:gl5457778; GSPDB:GN00174
C:Genetics:
A:Gene: PTS-EIIC
C:Superfamily: phosphotransferase system enzyme II factor II, phosphoenolpyruvate-de

Query Match 0.8%; Score 8; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 VPLOGLTG 93
|||||
DB 226 VPLOGLTG 233

RESULT 11
T11032
A:Title: dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - redhead mitochondrion
A:Species: mitochondrion Aythya americana (redhead)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
A:Accession: T11032
R:Mindell, D.P.; Sorenson, M.D.; Dimcheff, D.E.
Proc. Natl. Acad. Sci. U.S.A. 95, 10693-10697, 1998
A:Title: Multiple independent origins of mitochondrial gene order in birds.
A:Reference number: 217242
A:Accession: T11032
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-607 <JOH>
A:Cross-references: EMBL:AF090337; NID:g4887659; PID:g4887670; PIDN:AA32262.1
C:Genetics:
A:Gene: mitochondrion
A:Genetic code: SGI
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylati

Query Match 0.8%; Score 8; DB 2; Length 607;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 LTLTLATA 249
|||||
DB 411 LTLTLATA 418

RESULT 12
A81293
probable flagellar hook-associated protein Cj1466 [imported] - Campylobacter jejuni (str
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: A81293
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: A81293
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-608 <PAR>
A:CROSS-references: GB:ALJ39078; GB:ALJ11168; NID:g96968723; PIDN:CAB73889.1; PID:g696889
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: flgK; Cj1466

Query Match 0.8%; Score 8; DB 2; Length 608;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 GTGTAVES 768
DB 57 GTGTAVES 64

RESULT 13
A45638
immunodominant microneme protein Etp100 - Eimeria tenella
C:Species: Eimeria tenella
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A45638
R:Tomley, F.M.; Clarke, L.E.; Kawazoe, U.; Dijkema, R.; Kok, J.J.
Mol. Biochem. Parasitol. 49, 277-288, 1991
A:Title: Sequence of the gene encoding an immunodominant microneme protein of Eimeria te
A:Reference number: A45638; MUID:92131064; PMID:1775171
A:Accession: A45638
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-712 <TOM>
A:CROSS-references: GB:AF032905; GB:M73495; NID:g2707732; PIDN:AAD03350.1; PID:g2707733
A:Note: sequence extracted from NCBI backbone (NCBIN:77752, NCBI:P:77756)
C:Superfamily: thrombospondin type 1 repeat homology; von Willebrand factor type A repea
F:48-218/Domain: von Willebrand factor type A repeat homology <VWAL>
F:238-296/Domain: thrombospondin type 1 repeat homology <THR1>
F:309-371/Domain: thrombospondin type 1 repeat homology <THR2>
F:372-432/Domain: thrombospondin type 1 repeat homology <THR3>
F:433-493/Domain: thrombospondin type 1 repeat homology <THR4>
F:494-556/Domain: thrombospondin type 1 repeat homology <THR5>
F:560-610/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 0.8%; Score 8; DB 2; Length 712;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTC GGG 535
DB 446 CSRTC GGG 453

RESULT 14
T25061
hypothetical protein T21B6.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T25061
R:Cottage, A.
A:Reference number: Z19975

A:Accession: T25061
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-788 <WIL>
A:CROSS-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3
A:Experimental source: clone T21B6
C:Genetics:
A:Gene: CESP:T21B6.3
A:Map position: X
A:Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 0.8%; Score 8; DB 2; Length 788;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 851 CSASCGSG 858
DB 514 CSASCGSG 521

RESULT 15
T30525
alpha-mannosidase (EC 3.2.1.24) - Emericella nidulans
C:Species: Emericella nidulans, Aspergillus nidulans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T30525
R:Eades, C.J.; Gilbert, A.; Goodman, C.D.; Hintz, W.E.
Glycobiology 8, 17-33, 1998
A:Title: Identification and analysis of a class 2 alpha-mannosidase in Aspergillus ni
A:Reference number: Z20843; MUID:98119762; PMID:9451011
A:Accession: T30525
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1049 <EAD>
A:CROSS-references: EMBL:AF016850; NID:g2407175; PID:g2407176; PIDN:AAB70514.1
C:Genetics:
A:Note: msd
C:Superfamily: Saccharomyces alpha-mannosidase
C:Keywords: glycosidase; hydrolase

Query Match 0.8%; Score 8; DB 2; Length 1049;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 PLQGLTGG 94
DB 138 PLQGLTGG 145

Search completed: April 29, 2003, 17:25:03
Job time : 32 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:18:08 ; Search time 16 Seconds
(without alignments)

2462.657 Million cell updates/sec

Title: US-10-009-332-1

Perfect score: 950

Sequence: 1 MLLGLTLTAFAGRTAGGFE.....DQCNLHRKPOELDFCVLRPC 950

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	1.8	967	1	ATSL_HUMAN
2	17	1.8	967	1	ATSL_RAT
3	17	1.8	968	1	ATSL_MOUSE
4	13	1.4	630	1	ATSL_MOUSE
5	13	1.4	837	1	ATSL_HUMAN
6	13	1.4	1629	1	ATSL_HUMAN
7	12	1.3	905	1	ATSL_MOUSE
8	11	1.2	930	1	ATSL_HUMAN
9	11	1.2	930	1	ATSL_MOUSE
10	10	1.1	890	1	ATSL_HUMAN
11	10	1.1	1077	1	ATSL_HUMAN
12	10	1.1	1205	1	ATSL_BOVIN
13	10	1.1	1205	1	ATSL_HUMAN
14	10	1.1	1211	1	ATSL_HUMAN
15	9	0.9	860	1	ATSL_HUMAN
16	8	0.8	245	1	ATSL_BOVIN
17	8	0.8	408	1	IHH_CHICK
18	8	0.8	790	1	AD30_HUMAN
19	8	0.8	886	1	SM6B_MOUSE
20	8	0.8	887	1	SM6B_RAT
21	8	0.8	888	1	SM6B_HUMAN
22	8	0.8	997	1	ATSL_HUMAN
23	8	0.8	1593	1	ATSL_HUMAN
24	7	0.7	111	1	GLNB_RHILV
25	7	0.7	112	1	GLNB_AZOB
26	7	0.7	112	1	GLNB_RHILV
27	7	0.7	112	1	GLNB_RHILV
28	7	0.7	112	1	GLNB_RHILV
29	7	0.7	112	1	GLNB_RHILV
30	7	0.7	121	1	IAX1_WHEAT
31	7	0.7	131	1	CRB2_BACSU
32	7	0.7	143	1	RL15_METJA
33	7	0.7	172	1	IDE3_ERYLA

34	7	0.7	172	1	IDE3_ERYLA
35	7	0.7	172	1	IDE3_ERYLA
36	7	0.7	176	1	CFTR_MACMU
37	7	0.7	199	1	SP25_DROME
38	7	0.7	228	1	VG77_HSV11
39	7	0.7	244	1	FRDB_PROVU
40	7	0.7	253	1	YE24_MYCTU
41	7	0.7	274	1	PK1_NPVOP
42	7	0.7	302	1	CITG_KLEPN
43	7	0.7	311	1	PRMA_BACSU
44	7	0.7	317	1	GGH_RAT
45	7	0.7	321	1	K6PF_AQUAE

ALIGNMENTS

RESULT 1

ATSL_HUMAN STANDARD: PRT: 967 AA.

AC Q9UH8; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;

DT 30-MAY-2000 (Rel. 39, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).

GN ADAMTS1 OR METH1 OR KIAA1346.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Casas C., Fritchard M.A., Estivill X., Arbones M.L.;

RT "Cloning, characterization and mapping on human chromosome 21 of the orthologue of murine Adamts-1."

RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A., AND FUNCTION.

RC TISSUE=Heart;

RX MEDLINE=99367466; PubMed=10438512;

RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S., Lombardo M., Iruela-Arispe M.L.;

RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity."

RL J. Biol. Chem. 274:23349-23357(1999).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Endothelial cells;

RX MEDLINE=20247184; PubMed=10785405;

RA Glienke J., Schmitt A.O., Pillarsky C., Hinzmann B., Weiss B., Rosenthal A., Thierauch K.H.;

RT "Differential gene expression by endothelial cells in distinct angiogenic states."

RT Eur. J. Biochem. 267:2820-2830(2000).

RN [4]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=20181126; PubMed=10718198;

RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes.XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro."

RL DNA Res. 7:65-73(2000).

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE=20289799; PubMed=10830953;

RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,

RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordtsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramer J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.,
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [6]
RP SEQUENCE OF 418-967 FROM N.A.
RC : TISSUE-Melanoma;
RA Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR
CC ACTIVITY. ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1938-GLU-1-LEU-1939
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF170084; AAF15317.1; ALT.
DR EMBL; AF060152; AAD48080.1; ALT_INIT.
DR EMBL; AF207664; AAF23772.1; -
DR EMBL; AB037767; BAA92584.1; ALT_INIT.
DR EMBL; AP001697; BAA95502.1; -
DR EMBL; AL162080; CAB82413.1; -
DR MEROPS; M12.222; -
DR Genew; HGNC:217; ADAMTS1.
DR MW; 605174; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; zn_Mtpeptidse.
DR Pfam; PF000090; tsp.1; 6
DR Pfam; PF01421; Reprolysin; 2.
DR Pfam; PF01562; Pep_M12B_propep; 2.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS0215; ADAM_MPEPO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00427; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 49
FT PROPEP 50 252
FT CHAIN 253 967
FT SITE 198 198
FT METAL 401 401
FT ACT_SITE 402 402
FT METAL 405 405
FT METAL 411 411
FT DOMAIN 476 559
FT DOMAIN 560 616
TSP TYPE-1 1.

FT DOMAIN 617 724
FT DOMAIN 725 849
FT DOMAIN 850 908
FT DOMAIN 909 967
FT DOMAIN 943 846
FT CARBOHYD 547 547
FT CARBOHYD 720 720
FT CARBOHYD 764 764
FT CONFLICT 227 227
FT CONFLICT 468 468
FT CONFLICT 561 561
SQ SEQUENCE 967 AA; 105383 MW; C189389324741EDI CRC64;
Query Match 1.8%; Score 17; DB 1; Length 967;
Best Local Similarity 100.0%; Pred. No. 8.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 356 APTTAHELGHVFNMPHD 372
DB 396 APTTAHELGHVFNMPHD 412
|||||
RESULT 2
ATSL_RAT STANDARD; PRT; 967 AA.
AC Q9WUQ1; Q9ER11;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor [EC 3.4.24.-] (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RT "Induction of a disintegrin and metalloprotease with the
RT thrombospondin type I motif (ADAMTS).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 18-967 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Liver;
RX MEDLINE-20304099; PubMed-10847486;
RA Diamantis I., Luethi M., Hoesli M., Reichen J.;
RT "Cloning of the rat ADAMTS-1 gene and its down regulation in
RT endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CC CIRRHOTIC LIVER.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----


```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; AB001735; BAA24501.1; ALT.INT.
DR EMBL; D67076; BAA11088.1; ALT.FRAME.
DR MEROPS; M12.222; -.
DR MGD; MGI:109249; Adamts1.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep.M12B.propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; zn_Mtpeptdse.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep.M12B.propep; 1.
DR SMART; SM00209; TSPl; 3.
DR PROSITE; PS02015; ADAM_MPEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS50092; TSPl; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 48
FT PROPEP 49 253
FT CHAIN 254 968
FT SITE 206 206
FT METAL 402 402
FT ACT_SITE 403 403
FT METAL 406 406
FT METAL 412 412
FT DOMAIN 477 559
FT DOMAIN 560 617
FT DOMAIN 618 725
FT DOMAIN 726 850
FT DOMAIN 851 909
FT DOMAIN 910 968
FT DOMAIN 195 199
FT DOMAIN 548 548
FT CARBOHYD 721 721
FT CARBOHYD 765 765
FT CARBOHYD 783 783
FT CARBOHYD 946 946
FT MUTAGEN 403 403
FT CONFLICT 335 335
FT CONFLICT 425 425
FT SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;

Query Match 1.8%; Score 17; DB 1; Length 968;
Best Local Similarity 100.0%; Pred. No. 8.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 APTTAHELGHVFNMPHD 372
Db 397 APTTAHELGHVFNMPHD 413
|||||
RESULT 4
AT54_RAT STANDARD; PRT; 630 AA.
AC Q9ESP7; Q9ESP8; Q9ESP6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (Aggrecanase 1)
DE (Fragment).
DE GN ADAMTS4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RC STRAIN-Wistar; TISSUE=Brain;
RX MEDLINE=20415831; PubMed=10961658;
RA Satoh K., Suzuki N., Yokota H.;
RT "ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin
RT motifs) is transcriptionally induced in beta-amyloid treated rat
RT astrocytes.";
RL Neurosci. Lett. 289:177-180(2000).
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
CC site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: ONLY EXPRESSED IN BRAIN.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; AB042272; BAB16474.1; -.
CC EMBL; AB042271; BAB16473.1; -.
CC EMBL; AB042273; BAB16475.1; -.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSPl.
CC InterPro; IPR00130; zn_Mtpeptdse.
CC Pfam; PF00090; tsp_1; 2.
CC Pfam; PF01421; Reprolysin; 1.
CC SMART; SM00209; TSPl; 1.
CC PROSITE; PS02015; ADAM_MPEPRO; 1.
CC PROSITE; PS50092; TSPl; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Extracellular matrix.
KW NON_TER 1 1
FT PROPEP 1 5
FT CHAIN 6 630
FT METAL 154 154
FT ACT_SITE 155 155
FT METAL 158 158
FT METAL 164 164
FT DOMAIN 233 303
FT DOMAIN 316 367
FT DOMAIN 368 478
FT DOMAIN 479 630
FT DOMAIN 40 45
FT CARBOHYD 96 96
FT CARBOHYD 474 474
FT SEQUENCE 630 AA; 68384 MW; 63A428753167C7EF CRC64;

Query Match 1.4%; Score 13; DB 1; Length 630;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 DTAIFTRQDLGG 323
Db 104 DTAIFTRQDLGG 116
|||||
```

```

RESULT 5
AT94_HUMAN STANDARD; PRT; 837 AA.
AC 075173; Q9UN83;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-4 precursor (RC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1).
GN ADAMTS4 OR KIAA0688.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA Rockwell A., Tang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
RA Trzaskos J.M., Arner E.C.;
RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
RT family of proteins.";
RL Science 284:1664-1666(1999).
[3]
RP SEQUENCE FROM N.A.
RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;
RT "ADAMTS-4 genomic locus.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=10827174;
RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,
RA Burn T.C., Arner E.C.;
RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
RT aggrecan substrate recognition and cleavage.";
J. Biol. Chem. 275:25791-25797(2000).
CC -1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
CC ALZHEIMER'S DISEASE.
CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393
CC site.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
CC -1- INDUCTION: BY INTERLEUKIN-1.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; AB014588; BAA31663.1; -
CC EMBL; AF148213; AAD41494.1; -
CC EMBL; AY044847; AAL02262.1; -
CC MEROPS; M12.221; -
CC GENE; HGNC:220; ADAMTS4.
CC MIM; 603876; -
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_Mtpeptidse.
CC Pfam; PF00090; tsp_1; 1.
CC Pfam; PF01421; Reprolysin; 1.
CC SMART; SM00209; TSP1; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS00092; TSP1; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Extracellular matrix.
FT SIGNAL 1 51 POTENTIAL.
FT PROPEP 52 212
FT CHAIN 213 837
FT SITE 194 194
FT METAL 361 361
FT ACT_SITE 362 362
FT METAL 365 365
FT METAL 371 371
FT DOMAIN 437 519
FT DOMAIN 520 576
FT DOMAIN 577 685
FT DOMAIN 686 837
FT CARBOHYD 247 252
FT CONFLICT 68 68
FT CONFLICT 77 77
FT CONFLICT 626 626
FT CONFLICT 682 682
SQ SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;
Query Match 1.4%; Score 13; DB 1; Length 837;
Best Local Similarity 100.0%; Pred. No. 9.7e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 DTAILETRDLCG 323
DB 311 DTAILETRDLCG 323
(((((|||||))))
RESULT 6
ID AT94_HUMAN STANDARD; PRT; 1629 AA.
AC Q9P2N4; Q9NR29;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
GN ADAMTS9 OR KIAA1312.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Fetal;
RX MEDLINE=20396138; PubMed=10936055;

```


modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC DR EMBL; AF175282; AAF25805.1; -.
CC DR HSSP; P34179; LIAG.
CC DR MEROPS; M12.226; -.
CC DR MGD; MGI:1353468; Adamts8.
CC DR InterPro; IPR001762; Disintegrin.
CC DR InterPro; IPR001590; Reprolysin.
CC DR InterPro; IPR000884; TSPI.
CC DR InterPro; IPR000130; zn_Mtpeptdse.
CC DR Pfam; PF00090; tsp_1; 2.
CC DR Pfam; PF01421; Reprolysin; 1.
CC DR SMART; SM00209; TSPI_2.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
CC DR PROSITE; PS0215; ADAM_MEROPS; 1.
CC DR PROSITE; PS0092; TSPI; 1.
CC DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 228 BY SIMILARITY.
FT CHAIN 229 905 ADAMTS-8.
FT METAL 378 378 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 379 379 BY SIMILARITY.
FT METAL 382 382 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 453 541 DISINTEGRIN-LIKE.
FT DOMAIN 542 598 TSP TYPE-1 1.
FT DOMAIN 599 705 CYS-RICH.
FT DOMAIN 706 847 SPACER.
FT DOMAIN 848 905 TSP TYPE-1 2.
FT CARBOHYD 415 415 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 480 480 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 905 AA; 98879 MW; 124D4132B33A0CAE CRC64;

Query Match 1.3%; Score 12; DB 1; Length 905;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 603 WPKYSGVSPRD 614
DB 626 WPKYSGVSPRD 637
|||||||

* RESULT 8
AT55_HUMAN STANDARD; PRT; 930 AA.
AC Q9UNAO; Q9UKP2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (ADAM-TS 11).
GN ADAMTS5 OR ADMP2 OR ADAMTS11.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99367476; PubMed=10438522;
RA Abaszade I., Liu R.-O., Yang F., Rosenfeld S.A., Ross O.H.,
RA Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M.,
RA Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K.,
RA Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H.,
RA Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F.,
RA Arner E.C., Burn T.C.;

```

```

RT ADAMTS family.
RT J. Biol. Chem. 274:23443-23450(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reischwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Muroshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Reinhardt R., Yaspo M.-L.,
RA Lehrach H., Reinhardt R., Gardiner K., Nizetic D., Francis F.,
RT "The DNA sequence of human chromosome 21."
RT Nature 405:311-319(2000).
RN [3]
RP SEQUENCE OF 413-930 FROM N.A.
RX TISSUE=Fetal brain;
RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RA "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT zinc metalloproteases."
RL J. Biol. Chem. 274:25555-25563(1999).
CC -1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
CC PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393
CC site.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
CC BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
CC CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE, FROM AN
CC CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
CC ARTHRITIC PATIENT.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; AF142099; AAD49577.1; -.
CC EMBL; AP001698; BAA95504.1; -.
CC EMBL; AP001697; BAA95503.1; -.
CC EMBL; AF141293; AAF02493.1; -.
CC HSSP; Q9PW35; 1BUD.
CC MEROPS; M12.225; -.
CC Genew; HGNC:221; ADAMTS5.
CC MIM; 605007; -.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSPI.
CC InterPro; IPR000130; zn_Mtpeptdse.
CC Pfam; PF00090; tsp_1; 2.
CC Pfam; PF01421; Reprolysin; 1.

```


DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2)
DE (METH-8).
GN ADAMTS8 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT family of proteins with angio-inhibitory activity.";
RL J. Biol. Chem. 274:23349-23357(1999).
RN [2]
RP SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloproteinase of the ADAM-TS family located on
RT mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER
CC EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND
CC KIDNEY.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF060153; AAD48081.1; -;
DR EMBL; AF175283; AAF25806.1; -;
DR HSSP; P34179; 1IAG.
DR MEROPS; M12.226; -;
DR Genew; HGNC:224; ADAMTS8.
DR MIM; 605175; -;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptdse.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS50092; TSP1; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 214 BY SIMILARITY.
FT CHAIN 215 890 ADAMTS-8.
FT METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 365 365 BY SIMILARITY.

FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 439 526 DISINTEGRIN-LIKE.
FT DOMAIN 527 583 TSP TYPE-1 1.
FT DOMAIN 584 690 CYS-RICH.
FT DOMAIN 691 832 SPACER.
FT DOMAIN 833 890 TSP TYPE-1 2.
FT DOMAIN 202 205 POLY-PRO.
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 600 600 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 195 195 E -> R (IN REF. 2).
FT CONFLICT 413 440 YLFEILDGGHGDCLLDAPGAALPLPTGL -> FSGCHLOGW
FT IHPFKYLCCKVSELKCDLMP (IN REF. 2).
SQ SEQUENCE 890 AA; 96671 MW; 57D70EE03D5739D3 CRC64;

Query Match 1.1%; Score 10; DB 1; Length 890;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGGVQ 537
|||||
DB 539 CSRTCGGGVQ 548

RESULT 11
AT10_HUMAN
ID AT10_HUMAN STANDARD; PRT; 1077 AA.
AC Q9H324;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT thrombospondin type 1 repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF163762; AAG35563.1; -;
DR MEROPS; M12.235; -;
DR Genew; HGNC:13201; ADAMTS10.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptdse.

```

DR Pfam: PF00090; tsp_1; 5.
DR Pfam: PF01421; Repolysin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR SMART: SM00209; TSPl; 5.
DR PROSITE: PS00215; ADAM_MEPRO; 1.
DR PROSITE: PS00092; TSPl; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT NON_TER 1 1
FT PROPEP 1 207 BY SIMILARITY.
FT CHAIN 208 1077 ADAMTS-10.
FT METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 367 367 BY SIMILARITY.
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 CYS-RICH.
FT DOMAIN 680 802 SPACER.
FT DOMAIN 721 577 TSP TYPE-1 1.
FT DOMAIN 799 860 TSP TYPE-1 2.
FT DOMAIN 862 918 TSP TYPE-1 3.
FT DOMAIN 922 976 TSP TYPE-1 4.
FT DOMAIN 981 1031 TSP TYPE-1 5.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 769 769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE1DCBFB587 CRC64;

Query Match 1.1%; Score 10; DB 1; Length 1077;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 218 RYVETLVVAD 227
Dy 213 RYVETLVVAD 222
|||||

RESULT 12
AT52_BOVIN STANDARD; PRT; 1205 AA.
AC P79331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI).
GN ADAMTS2 OR NPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RX MEDLINE=97255960; PubMed=9122202;
RA Colige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,
RA Lapierre C.M.;
RT "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT new member of the superfamily of zinc-metalloproteinases with binding
RT sites for cells and other matrix components";
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=95348096; PubMed=7622483;
RA Colige A., Beschin A., Samyn B., Goebels Y., Van Beeumen J.,

```

```

RA Nusgens B.V., Lapierre C.M.;
RT "Characterization and partial amino acid sequencing of a 107-kDa
RT procollagen I N-proteinase purified by affinity chromatography on
RT immobilized type XIV collagen.";
RL J. Biol. Chem. 270:16724-16730(1995).
CC -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC COLLAGEN BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains
CC at Ala-1-Gln.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC COLLAGEN TYPE XIV.
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X96389; CAA65253.1; -.
DR MEROPS; M12.301; .
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 4.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPl; 4.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPl; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28
FT PROPEP 29 253 BY SIMILARITY.
FT CHAIN 254 1205 ADAMTS-2.
FT METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 403 403 BY SIMILARITY.
FT METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 474 554 DISINTEGRIN-LIKE.
FT DOMAIN 555 611 TSP TYPE-1 1.
FT DOMAIN 612 716 CYS-RICH.
FT DOMAIN 717 845 SPACER.
FT DOMAIN 846 905 TSP TYPE-1 2.
FT DOMAIN 906 968 TSP TYPE-1 3.

```

FT DOMAIN 969 1024 TSP TYPE-1 4.
 FT SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).
 FT DOMAIN 31 35 POLY-ALA.
 FT DOMAIN 177 180 POLY-GLU.
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 943 943 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 987 987 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;

 Query Match 1.1%; Score 10; DB 1; Length 1205;
 Best Local Similarity 100.0%; Pred. No. 0.17;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 670 DKCGVCGGDN 679
 Db 704 DKCGVCGGDN 713
 |||||

 RESULT 13
 AT53_HUMAN
 ID AT53_HUMAN STANDARD; PRT; 1205 AA.
 AC O15072; Q9BXZ8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 3) (ADAM-TS 3) (Procollagen II
 DE amino-propeptide processing enzyme) (Procollagen II N-proteinase) (PC
 DE II-NP).
 GN ADAMTS3 OR KTA0366.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-227 FROM N.A.
 RX MEDLINE=21402912; PubMed=11408482;
 RA Fernandes R.J., Hirohata S., Engle J.M., Colige A., Cohn D.H.,
 RA Eyre D.R., Apte S.S.;
 RT "Procollagen II amino propeptide processing by ADAMTS-3. Insights on
 RT dermatosparaxis.";
 RL J. Biol. Chem. 276:31502-31509(2001).
 RN [2]
 RP SEQUENCE OF 5-1205 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 CC -!- FUNCTION: Cleaves the propeptides of type II collagen prior to
 CC fibril assembly. Does not act on types I and III collagens.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -!- TISSUE SPECIFICITY: Found in cartilage and skin.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF247668; AAK28400.1; -;
 DR EMBL; AB002364; BAA20821.1; -;
 DR MEROPS; M12.220; -;
 DR Genew; HGNC:219; ADAMTS3.
 DR MTM; 605011; -;
 DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR002870; Pep_M12B_propep.
 DR InterPro; IPR001590; Reprolysin.
 DR InterPro; IPR000884; TSPI.
 DR InterPro; IPR000130; Zn_MTpeptdse.
 DR Pfam; PF00090; tsp_1; 4.
 DR Pfam; PF01421; Reprolysin; 1.
 DR Pfam; PF01562; Pep_M12B_propep; 1.
 DR SMART; SM00209; TSPI; 4.
 DR PROSITE; PS00215; ADAM_MEPRO; 1.
 DR PROSITE; PS0092; TSPI; 2.
 DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 KW Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 249 BY SIMILARITY.
 FT CHAIN 250 1205 ADAMTS-3.
 FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 399 399 BY SIMILARITY.
 FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 470 550 DISINTEGRIN-LIKE.
 FT DOMAIN 551 607 TSP TYPE-1 1.
 FT DOMAIN 608 712 CYS-RICH.
 FT DOMAIN 713 844 SPACER.
 FT DOMAIN 845 902 TSP TYPE-1 2.
 FT DOMAIN 903 965 TSP TYPE-1 3.
 FT DOMAIN 966 1017 TSP TYPE-1 4.
 FT DOMAIN 246 249 POLY-ARG.
 FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1205 AA; 135570 MW; EB07B286FC85FB87 CRC64;

 Query Match 1.1%; Score 10; DB 1; Length 1205;
 Best Local Similarity 100.0%; Pred. No. 0.17;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 670 DKCGVCGGDN 679
 Db 700 DKCGVCGGDN 709
 |||||

 RESULT 14
 AT52_HUMAN
 ID AT52_HUMAN STANDARD; PRT; 1211 AA.
 AC Q95450;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-OCT-2001 (Rel. 41, Last annotation update)
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
 DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
 DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
 DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI)
 DE (Procollagen I/II amino-propeptide processing enzyme).

GN ADAMTS2 OR PCINP OR PCPNI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.
 RC TISSUE=skin;
 RA MEDLINE=99347935; PubMed=10417273;
 RA Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,
 RA Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
 RA Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.;
 RA "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis
 are caused by mutations in the procollagen I N-proteinase gene.";
 RL Am. J. Hum. Genet. 65:308-317(1999).
 CC -1- FUNCTION: Cleaves the propeptides of type I and II collagen prior
 CC to fibril assembly. Does not act on type III collagen. May also
 CC play a role in development that is independent of its role in
 CC collagen biosynthesis.
 CC -1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
 CC alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains
 CC at Ala-1-Gln.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
 CC COLLAGEN TYPE XIV (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-
 CC PROCOLLAGEN PEPTIDASE ACTIVITY.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON
 CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -1- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos
 CC syndrome type VIIC (EDS-VIIC), a recessively inherited connective-
 CC tissue disorder characterized clinically by severe skin fragility
 CC and joint hypermobility and biochemically by the presence in skin
 CC of procollagen incompletely processed at the N-terminus.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ003125; CAA05880.1; -
 DR MEROPS; M12.301; -
 DR Genew; HGNC:218; ADAMTS2.
 DR MIN; 604539; -
 DR MIN; 225410; -
 DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR002870; Pep_M12B_propep.
 DR InterPro; IPR001590; Reprolysin.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR000130; zn_MTPeptidse.
 DR Pfam; PF00090; tsp-1; 4.
 DR Pfam; PF01421; Reprolysin; 1.
 DR Pfam; PF01562; Pep_M12B_propep; 1.
 DR SMART; SM00209; TSP1; 4.
 DR PROSITE; PS02015; ADAM_MEPPO; 1.
 DR PROSITE; PS00092; TSP1; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR HydroLase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW

KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
 KW Alternative splicing; Ehlers-Danlos syndrome.
 FT SIGNAL 1 29
 FT PROPEP 30 253
 FT CHAIN 1211
 FT CHAIN 408 408
 FT METAL 409 409
 FT ACT_SITE 412 412
 FT METAL 418 418
 FT METAL 480 560
 FT DOMAIN 561 617
 FT DOMAIN 618 722
 FT SITE 691 693
 FT DOMAIN 723 851
 FT DOMAIN 852 911
 FT DOMAIN 912 974
 FT DOMAIN 975 1030
 FT DOMAIN 40 43
 FT DOMAIN 185 188
 FT CARBOHYD 112 112
 FT CARBOHYD 251 251
 FT CARBOHYD 949 949
 FT CARBOHYD 993 993
 FT CARBOHYD 1031 1031
 FT CARBOHYD 1098 1098
 FT CARBOHYD 1145 1145
 FT CARBOHYD 1150 1150
 FT VARSPLIC 544 566
 FT VARSPLIC 567 1211
 FT SEQUENCE 1211 AA; 134722 MW; BECEF25C33CAD2D CRC64;
 SQ

Query Match 1.1%; Score 10; DB 1; Length 1211;
 Best Local Similarity 100.0%; Pred. No. 0.17;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 DKCGVCGGDN 679
 Db 710 DKCGVCGGDN 719

RESULT 15

AT56.HUMAN
 ID AT56.HUMAN STANDARD; PRT; 860 AA.
 AC Q9URP5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-6 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 6) (ADAM-TS 6) (ADAM-TS6).
 GN ADAMTS6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99395124; PubMed=10464288;
 RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
 RA "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
 RT Zinc Metalloproteases";
 RL J. Biol. Chem. 274:25555-25563(1999).
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN PLACENTA AND BARELY
 CC DETECTABLE IN A NUMBER OF OTHER TISSUES.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

```
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF140674; AAD56357.1; .
DR HSSP; P34179; LIAG.
DR MEROPS; M12.230; .
DR Genew; HGNC:222; ADAMTS6.
DR MIM; 605008; .
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; zn_MTpeptdse.
DR Pfam; PF00090; tsp.1; 1.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS00215; ADAM_MPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00092; TSPI; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21
FT PROPEP 22 244
FT CHAIN 245 860
FT METAL 403 403
FT ACT_SITE 404 404
FT METAL 407 407
FT METAL 413 413
FT DOMAIN 453 509
FT DOMAIN 510 566
FT DOMAIN 567 668
FT DOMAIN 669 795
FT DOMAIN 796 852
FT DOMAIN 68 71
FT DOMAIN 662 665
FT CARBOHYD 99 99
FT CARBOHYD 172 172
FT CARBOHYD 222 222
FT CARBOHYD 234 234
FT CARBOHYD 676 676
FT CARBOHYD 843 843
SQ SEQUENCE 860 AA; 97098 MW; E57213015DECB2C5 CRC64;

Query Match 0.9%; Score 9; DB 1; Length 860;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGV 536
Db- 522 CSRTCGGV 530
|||||||
```

Search completed: April 29, 2003, 17:23:57
Job time : 19 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 22.6233 Seconds
(without alignments)
3378.970 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_583

Perfect score: 2042
Sequence: 1 FVSIPRYVETLVVADESMVK.....SCNLEPCPSSASGKSFREQ 371

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phase:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-rvirus:*
- 16: sp-bacteriap:*
- 17: sp-archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2042	100.0	950	4 Q8TE58	Q8TE58 homo sapien
2	923.5	45.2	2165	5 Q19791	Q19791 caenorhabdi
3	796	39.0	269	6 Q9GL54	Q9GL54 oryctolagus
4	796	39.0	1054	5 Q9W493	Q9W493 drosophila
5	703	34.4	1688	5 Q8SX80	Q8SX80 drosophila
6	670.5	32.8	1229	5 Q9VF61	Q9VF61 drosophila
7	654	32.0	1072	4 Q8TE57	Q8TE57 homo sapien
8	643.5	31.5	1095	4 Q8TE56	Q8TE56 homo sapien
9	638.5	31.3	1081	4 Q8TE60	Q8TE60 homo sapien
10	617.5	30.2	1223	4 Q8WX58	Q8WX58 homo sapien
11	617.5	30.2	1223	4 Q8TE55	Q8TE55 homo sapien
12	616	30.2	1159	4 Q8TEY8	Q8TEY8 homo sapien
13	606	29.7	1207	4 Q8TE59	Q8TE59 homo sapien
14	558	27.3	192	6 Q95N24	Q95N24 equus caball
15	540	26.4	1427	4 Q96L37	Q96L37 homo sapien
16	518	25.4	187	6 Q95N23	Q95N23 equus caball

17	396.5	19.4	1062	5 Q19204	Q19204 caenorhabdi
18	393	19.2	872	5 Q22580	Q22580 caenorhabdi
19	377.5	18.5	1444	5 Q17591	Q17591 caenorhabdi
20	361.5	17.7	1091	5 Q9W126	Q9W126 drosophila
21	357.5	17.5	790	5 Q8T458	Q8T458 drosophila
22	303	14.8	108	6 Q9GLK6	Q9GLK6 oryctolagus
23	298	14.6	117	11 Q8VHK4	Q8VHK4 mus musculu
24	237.5	11.6	364	4 Q9UGQ1	Q9UGQ1 homo sapien
25	231.5	11.3	509	5 Q9NDL4	Q9NDL4 caenorhabdi
26	231.5	11.3	509	5 Q20930	Q20930 caenorhabdi
27	231.5	11.3	899	13 Q8UVF1	Q8UVF1 coturnix co
28	226	11.1	610	13 Q93523	Q93523 bothrops ja
29	226	11.1	610	13 Q8QG88	Q8QG88 bothrops in
30	224.5	11.0	477	13 Q98SP2	Q98SP2 bothrops ja
31	224	11.0	922	13 Q8UVF2	Q8UVF2 coturnix co
32	223.5	10.9	609	13 Q9W6M5	Q9W6M5 agkistrodon
33	221	10.8	482	13 Q9PVK9	Q9PVK9 agkistrodon
34	220.5	10.8	465	13 Q9TAX7	Q9TAX7 agkistrodon
35	220.5	10.8	479	13 Q9PWJ0	Q9PWJ0 agkistrodon
36	216.5	10.7	419	13 Q92043	Q92043 crotalus at
37	216.5	10.6	476	13 Q9YI19	Q9YI19 agkistrodon
38	214.5	10.5	479	13 Q9PW78	Q9PW78 agkistrodon
39	213	10.4	483	13 Q9IAB0	Q9IAB0 agkistrodon
40	213	10.4	549	13 Q90500	Q90500 echis pyram
41	212	10.4	4123	4 Q75851	Q75851 homo sapien
42	210.5	10.3	612	13 Q8UVG0	Q8UVG0 bothrops er
43	210	10.3	610	13 Q9YI20	Q9YI20 agkistrodon
44	209.5	10.3	617	13 Q90499	Q90499 echis pyram
45	208.5	10.2	609	13 Q90282	Q90282 crotalus at

ALIGNMENTS

RESULT 1

Q8TE58	PRELIMINARY;	PRT;	950 AA.
AC Q8TE58;			
DT 01-JUN-2002 (TREMREL. 21, Created)			
DT 01-JUN-2002 (TREMREL. 21, Last sequence update)			
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)			
DE Metalloprotease disintegrin 15 with thrombospondin domains.			
GN ADAMTS15.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=21856482; PubMed=11867212;			
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,			
RA Lopez-Otin C.;			
RT seven novel human ADAMTSs, a family of metalloproteinases with			
RT disintegrin and thrombospondin-1 domains.*;			
RL Gene 283:49-62(2002).			
DR EMBL; AJ315733; CAC86014.1; -			
KW Integrin; Protease.			
SQ SEQUENCE 950 AA; 103286 MW; 5DFBE18285CCCC3B CRC64;			

Query Match 100.0%; Score 2042; DB 4; Length 950;
Best Local Similarity 100.0%; Pred. No. 1.1e-195;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	FVSIPRYVETLVVADESMVKFHGADLEHYLLTLAARLYRHPSILNFINVVVKVLL 60
Db	213	FVSIPRYVETLVVADESMVKFHGADLEHYLLTLAARLYRHPSILNFINVVVKVLL 272
QY	61	RDRDGPVKVGNAAATLNFCAWOKKLNKVSQKPEYWDTAILETRDLCGATTCDFLGM 120
Db	273	RDRDGPVKVGNAAATLNRNCAWOKKLNKVSQKPEYWDTAILETRDLCGATTCDFLGM 332
QY	121	ADVGTMCDFKRSCVIEDDGLPFAFTTAHELGHVFNMPHDNVKCEVFGKLRANHMSP 180

```
Db 333 ADYGTWCDPKRSVTEDDGLPSAFTTAHELGHVFNPHDNVKNVCEYFGKLRANHMSP 392
|||
QY 181 TLQIDRANPWSACSAAITDFLDSHGDCILLDQSPKPSLPEDLPAGASYTLSSQOCELAF 240
|||
Db 393 TLQIDRANPWSACSAAITDFLDSHGDCILLDQSPKPSLPEDLPAGASYTLSSQOCELAF 452
|||
QY 241 GVSGKPCPYMYCTKLWCTGKAGQKVCQTRHFPWADGTSOGEGLKLGACVERHNLNK 300
|||
Db 453 GVSGKPCPYMYCTKLWCTGKAGQKVCQTRHFPWADGTSOGEGLKLGACVERHNLNK 512
|||
QY 301 HRYDGSWAKWDYGPSCSRTCCGGGVOLARQCTNPTPANGGKYCEGVRYKYRSCNLEPCPS 360
|||
Db 513 HRYDGSWAKWDYGPSCSRTCCGGGVOLARQCTNPTPANGGKYCEGVRYKYRSCNLEPCPS 572
|||
QY 361 SASGKSFRREQ 371
|||
Db 573 SASGKSFRREQ 583
|||

RESULT 2
Q19791 PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP Gajadasy S.;
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RX MDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons N., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.;
RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1;
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1;
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSSP; P15167; 1DTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00209; TSP1; 18.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AA9C4888 CRC64;

Query Match 45.2%; Score 923.5; DB 5; Length 2165;
Best Local Similarity 45.4%; Pred. No. 4.3e-83;
Matches 177; Conservative 60; Mismatches 126; Indels 27; Gaps 7;
```

```
QY 7 YVELTVVADESVMKPHGADLEHYLLTLLTAARLYRHPHSILNPINIVVVKVLLLRDRDSG 66
|||
Db 281 YVEVLVAADTKMVEYHGRSLDEYVLTFTSTVASIYRHQSLRASINVVVVKLVLTENAG 340
|||
QY 67 PKVTGNAATLNFCAWQKKLNKVDKHPYWDTAILFTRODLGAT--TCDTLGMADVGT 125
|||
Db 341 PRITQNAOQTLODFCRWQYYNDPDDSSVQHHDVAILLTRKIDICKRSQCKDTLGLAELGT 400
|||
QY 126 MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEV-----FGKL 172
|||
Db 401 MCDMQSKCAIIEDNGLSAAFTIAHELGHVFSIPHDERKCKSTYMPVKNVKCFQSTKEDKT 460
|||
QY 173 RAN---HMSPPLIQIDRANPWSACSAAITDFLDSHG---DCLLDQSPKPSLP 223
|||
Db 461 OFQNNPHIMAPTLEYNTHPWSWSPCSAGMLERFLENNGQTOCLFDQPVERRYEDVFVR 520
|||
QY 224 DLPGASYTLSSQOCELAFGVGSKPCPYMYCTKLWCTGKAGQKVCQTRHFPWADGTSOGE 283
|||
Db 521 DEPGKKYDAHQCKKFVGPASSELCPYMPTCRRLWCATFYGSQMGCRTOHMPWADGTPCDE 580
|||
QY 284 GK--ICLKAGACVERHNLNKRVDSWAKWDYGPSCSRTCCGGGVOLARQCTNPTPANGGK 341
|||
Db 581 SRSMFCHGACVRLAPESLTIKIDQMGDMRWGECSTCCGGGVQKGLRDCDSPKPRNGGK 640
|||
QY 342 YCEGVRYKYRSCNLEPCPSASGKSFRREQ 371
|||
Db 641 YCVGQREYRSCNTOECPWDT--QPYREVQ 668
|||

RESULT 3
Q9GL54 PRELIMINARY; PRT; 269 AA.
ID Q9GL54;
AC Q9GL54;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Aggrecanase-2 (Fragment).
GN ADAMTS-11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Coad D.L., Coad M.E.;
RT "Molecular cloning of lapine aggrecanase-2 (ADAMTS-11) from articular
RT chondrocytes.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF317415; AAG33062.1; -.
DR HSSP; Q9FW35; 1BUD.
DR MEROPS; M12.225; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 269
SQ SEQUENCE 269 AA; 29193 MW; 97A1CA80B33452FA CRC64;

Query Match 39.0%; Score 796; DB 6; Length 269;
Best Local Similarity 53.8%; Pred. No. 1.7e-71;
Matches 140; Conservative 37; Mismatches 73; Indels 10; Gaps 3;
```

```
QY 117 TLGMADVGTMCDPKRSVTEDDGLPSAFTTAHELGHVFNPHDNVKNVCEYFGKLRANH 176
|||
Db 1 TLGMADVGTICSPERSCAVIEDDGLHAAFTVAHEITGLLGLSHDSDSKFCEENFGSTEDKR 60
|||
```



```
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Metalloprotease disintegrin 16 with thrombospondin type I motif.
GN ADAMTS16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315734; CAC86015.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1072 AA; 119491 MW; 081EEFF78F47D061 CRC64;

Query Match 32.0%; Score 654; DB 4; Length 1072;
Best Local Similarity 39.5%; Pred. No. 1.8e-56;
Matches 149; Conservative 49; Mismatches 149; Indels 30; Gaps 12;

QY 8 VETLVVADESVMVKFHG-ADLEHYLLTLTATAARLYRHPSILNPINIVVVKVLLLRDRSG 66
DB 292 VETLVVDDKMKQNHGHENITVYVLTILNMVSALFKDGTIGGNINIAIVGLLEDEQPG 351
QY 67 PKVTGNAALTLRNFCAMQKKL-NKVSQKHPEYWDTAILETRQDLGG-ATTCDTLGMAV 123
DB 352 LVISHHADHTLSSFCQWQGLMGKDGTRH----DHAILLTGIDCSKNWNEPCDTLGFAP 407
QY 124 GTMCDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDNPV-KVCEVFGLKLANHMSPTL 182
DB 408 SGMCSKYRSCITNEDTGLGLAFTIAHESGHNFMIHDGEGNCKKSEQ-----NIMSP 462
QY 183 IOIDRANFWSACSAAIITDFLDGSHGDCLLDQPSKPI---SLPEDLPGASYLSQOCELA 239
DB 463 AGRNQVFWSPCSRSQYHLKFLSTAQAICLADOP-KPVKEYKYPEKLPGELYDANTOCKW 521
QY 240 FGVGSKPCPY---MQCYCKLWC--TGKAGQMVQCOTRHFHPWADGTSCEGKLCILKGACVE 294
DB 522 FGEKAKLMLDPKDKICAKLWCHRIGRK-----CETKEPFAAEGTICGDMWCRGGQCVK 576
QY 295 RHNLNKHVRDGSWAKWDPYGPCSRCTCGGVQVLRQCTNPTNPANGKGYCEGVRVRYRSCN 354
DB 577 YGDEGPKPTHGHWSWSSWSPCSRTCGGVSHRSRLCTNPKPSHGKFCGEGSTRTLKLCN 636
QY 355 LEPCCPSSASGKSFREQ 371
DB 637 SQKCPRDS--VDFRAAQ 651

RESULT 8
Q8TE56 PRELIMINARY; PRT; 1095 AA.
AC Q8TE56;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Metalloprotease disintegrin 17, with thrombospondin domains.
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315734; CAC86015.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1072 AA; 119491 MW; 081EEFF78F47D061 CRC64;

Query Match 31.5%; Score 643.5; DB 4; Length 1095;
Best Local Similarity 36.7%; Pred. No. 2.1e-55;
Matches 142; Conservative 64; Mismatches 146; Indels 35; Gaps 12;

QY 8 VETLVVADESVMVKFHGAD-LEHYLLTLTATAARLYRHPSILNPINIVVVKVLLLRDRSG 66
DB 234 VETLVVADADVMQYHGAQAQRFILTMVMVYMFQHQSLGKIKINIQVTKVLLQRPAK 293
QY 67 PKVTGNAALTLRNFCAMQKK-----LNKV--SDKHPEYWDTAILETRQDLG--GAT 113
DB 294 LSIHHGERSLESFCHQWNEEYGGARYLGNQVPGKDDPPLVDAAVFVTRTDFCVHKDE 353
QY 114 TCDTLGMADVGMCDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDNPV-KVCEVFGLK 173
DB 354 PCDTVGIAYLGGVCSAKRKCVLAEDNGLNLAFTTIAHELGNLGNMNH-----DDHSS 408
QY 174 ANHMSPTLIQIDRANP-----WSACSAAIITDFLDGSHGDCLL---DQPSKPISLP 226
DB 409 RSHIMGEWVK--GRNPSDLSSSCSRDDLENFLKSKVSTCLLVTPRSHQHTVRLPHKL 466
QY 227 GASYTILSQCELAFGVSGKPCPYMQY--CTKLWCTGKAGQMVQCOTRHFHPWADGTSCEG 284
DB 467 GMHYSANEQCIILFGNNAIFCRNMEHLMAAGLWCL--VEGDTSCKTLPDPLDGTCEGAD 524
QY 285 KLCILKAGCVERHNLNKHVRDGSWAKWDPYGPCSRCTCGGVQVLRQCTNPTNPANGKYCE 344
DB 525 KWRAGECVSKTPIPEH-VGDWSPWGAWSMCSRTCTGTFARFQRKCDNPPGPGGTHCP 583
QY 345 GVRVYKRSCLNLEPCPSSASGKSFREQ 371
DB 584 GASVEHAVCENLPCPKGL--PSFRDQ 608

RESULT 9
Q8TE60 PRELIMINARY; PRT; 1081 AA.
AC Q8TE60;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Metalloprotease disintegrin 21, last annotation update)
DE ADAMTS18 protein.
GN ADAMTS18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ311903; CAC83612.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1081 AA; 119655 MW; 0438BF645676461E CRC64;

Query Match 31.3%; Score 638.5; DB 4; Length 1081;
Best Local Similarity 38.0%; Pred. No. 6.6e-55;
Matches 143; Conservative 51; Mismatches 153; Indels 29; Gaps 12;

QY 8 VETLVVADESVMVKFHG-ADLEHYLLTLTATAARLYRHPSILNPINIVVVKVLLLRDRSG 66
DB 295 VETLVVADKRWVKEHGKGNVTYILTVMKVSGL-DFKDGITGSDINVVVSVLLEQEPG 353
```

Qy	67	PKVTGNAALTURNFCAMQKKL-NKVS DKHPHYMDTAILFTRODLQG--ATTCDTLGMADV	123
Db	354	LLINHHADOSUNSFQMSQALIGNGRRH---DHALLLTGFDICSWKNNEPCDTLGFAPT	409
Qy	124	GTMTCDPKRCSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEEVFGKLRANHHMSPTL	182
Db	410	SGMCSKYRSCSTINEDTGLGLAFTTAHESGHFNFGMIHDGEGNPCRAEG-----NIN	464
Qy	183	IQIDRANPWASCSAAIITDLDLSCHGDCILLDQPSK--PTSLPEDLPASVYTSQQCELAF	240
Db	465	TGNNGVFSWCSRSQYLKKFLSTPQAGLQDEPKAQGYKYPDKLPQIYDADTQCKWQF	524
Qy	241	GVGSKPCP--YMOYCFKLWC--TGRAKGQWVCQTRHEFWADGTSCGEGKCLKLGCACVER	295
Db	525	GAKAKLSLGFVRKDICKSLNCHRVGHR-----CETFMFAAEGTVCGLSMWCQGQCVKF	579
Qy	296	HNLNKHRVDGSAWKDPYGPCSTCGGGVOLARRQCTNPTTPANGGKYCEGVVRKYRSCNL	355
Db	580	GELGPRPIHGOWSAWSKWSCSRTCGGGVKQFQERHCNNPKPYGGIFPCGSSRSIYQLCNI	639
Qy	356	EPCPSSASGSKSFREQ	371
Db	640	NPC--NENSLDFRAQQ	653

RESULT 10	
Q8WX58	
ID	Q8WX58 PRELIMINARY; PRT; 1223 AA.
AC	Q8WX58;
DT	01-MAR-2002 (TrEMBLrel. 20, Created)
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	A disintegrin-like and metalloprotease with thrombospondin type 1 motif 14 precursor.
DE	GN ADAMTS14.
GN	ADAMTS14.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC	NCHI_taxid=9606;
OR	[1]
RN	SEQUENCE FROM N.A.
RP	MEDLINE=21638061; PubMed=11779638;
RX	Bolz H., Ramirez A., von Bredelow B., Kubisch C.;
RT	"Characterization of ADAMTS14, a novel member of the ADAMTS
RT	metalloprotease family.";
RL	Biochim. Biophys. Acta 1522:221-225(2001).
RL	EMBL; AF358666; AAL40229.1; -.
DR	InterPro; IPR0012870; Pep_M12B_propep.
DR	InterPro; IPR001590; Reprolysin.
DR	InterPro; IPR000884; TSP1.
DR	Pfam; PF01562; Pep_M12B_propep; 1.
DR	Pfam; PF01421; Reprolysin; 1.
DR	Pfam; PF00090; tsp-1; 4.
DR	SMART; SM00209; TSP1; 4.
DR	PROSITE; PS50215; ADAM_MPEPRO; 1.
DR	PROSITE; PS50092; TSP1; 1.
KW	Signal; Integrin; Protease; Metalloprotease.
FT	SIGNAL 1 22 POTENTIAL.
FT	SEQUENCE 1223 AA; 133871 MW; 36394ACD92F170F CRC64;

Query Match 30.28; Score 617.5; DB 4; Length 1223;
Best Local Similarity 35.98; Pred. No. 9.9e-53;
Matches 134; Conservative 59; Mismatches 155; Indels 25; Gaps 10;

[illegible]

QY	126	MCDPKRSCSVIEDGGLPSAFTTAHELGHVFNMPHD---NVKYCEEVFVKLRANHMMSPTL	182
Db	375	MCHPLRSCLNHEDFSSAFVIAHETGHVLGMEDHGQGNGCADETSLGS-----VMAPLV	429
QY	103	IQIDRNPWSACSAAIITDFLDSGHGDCLLDQPSKPI-SLPEDLPASYYTLSSQOCELAFG	241
Db	430	QAAPHRFHWSCRCSKLELSRYLPS--YDCILDDPDPAPWQPPELPGINYSMDQCREFDG	487
QY	242	VGSKPC---PYMYCTKLWCTGRKAGOMVCOTRRHPWADGTSCGGKCLKKGACVERHNL	298
Db	488	SGYTQLAFRTFEPCKOLMCGS-HPDNYPYCKTKGPDLDGTGECAPGKCFCFHGCIWKSPE	546
QY	299	NKHVRDGSWAQWDYPGPCSTCGGVQLARRQCTNTPANGKKYCEGYRVKYRSCNLEPC	358
Db	547	QTYGODGWSSWTKFSGCSRSCGGVYRSRSCNNPSPAYGGRPCPLGPMFFFYQVCNSEEC	606
QY	359	PSSASGKSFRREQ	371
Db	607	PGTY--EDFRAQ	617
RESULT 11			
QBTE55 PRELIMINARY; PRT; 1223 AA.			
ID	QBTE55	PRELIMINARY;	PRT; 1223 AA.
AC	QBTE55;		
DT	01-JUN-2002 {T-EMBLrel. 21, Created}		
DT	01-JUN-2002 {T-EMBLrel. 21, Last sequence update}		
DT	01-JUN-2002 {T-EMBLrel. 21, Last annotation update}		
GN	Metalloprotease-disintegrin protease.		
DN	ADAMTS14.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=FETAL LUNG;		
RX	MEDLINE=21856482; PubMed=11867212;		
RA	Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V., Lopez-Otin C.;		
RT	"Cloning, expression analysis, and structural characterization of seven novel human ADAMTSs, a family of metalloproteinases with disintegrin and thrombospondin-1 domains.";		
RL	Gene 283;49-62(2002)		
DR	EMBL; AJ345098; CAC87943.1; -.		
KW	Integrin; Protease.		
SQ	SEQUENCE 1223 AA; 133925 MW; D585B6593977ED15 CRC64;		

Query Match 30.2%; Score 617.5; DB 4; Length 1223;
Best Local Similarity 35.9%; Pred. No. 9.9e-53;
Matches 134; Conservative 59; Mismatches 155; Indels 25; Gaps 10

Qy	8	VETLVVADESMVKFPHGAD - LEHYLLTLLTAARLYRHPISILNPINIVVKKVLLLRDRDGS	66
Db	261	IEVLVVYDDSVRFGRKEHQVNTVLTLMNIVDEYIHDESLGVHINIALVRLIMVGVRQSL	320
Qy	67	PKV - TGNAAUULTRNFCAWOKKLANKSDKHPEYWDTAILLTRQDLCGATTCDDTLGMDVGT	125
Db	321	SLIERGNPSRLSEQVCEWAHSQOQDPSHAHHDHVVFLTRQDF - - - - -GPSGYAPVTG	374
Qy	126	MCDPKRSCSVIEDDGLPSATTTAHELGHVNMPHD - - - NVKCEEYFGKLRAHHMMSPTL	182
Db	375	MCHPLRSCALNHEDGFSFAFVIAHETGVLGMHDGQGNCADETSIGS - - - - -VMAPIV	429
Qy	183	IQIDRANPWSACSAALITDPLDSGHGDCILLDQPSKPI - SLPELDPCASVYLSQOCELAFG	241
Db	430	QAAFHREHWSRCSKLELSRYLPS - YDCLLDQDFDPAWPQPELPFOINYSMDQCFRDFG	487
Qy	242	VGSKPC - - - PYMOYCTKWLCTGAKQOMVCOTRHPFWADGTSGGEGKCLKKACQVERHNL	298
Db	488	SGYQTCLAFPTFPCKOLWCS - HPDNPYFCYCTKKGPPLDGTGECAPQGWCFKGHC1WKSP	546
Qy	299	NKHRVDGSNAKWDPYGPCSRCTCGGQVLARRQCTNPTPANGGKYCGEVRVYRSCNLEPC	358

```

Db 547 QTYGQDGGSSWTKFGSCRSOGGVRSSRCNNPSPAYGGRPCLGPMFEYQVCNSEEC 606
QY 359 PSSASGKSPREQ 371
Db 607 PGTY--EDFRAQ 617

RESULT 12
Q8TE58 PRELIMINARY; PRT; 1159 AA.
AC Q8TE58;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ADAMTS14.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21839041; PubMed=11741898;
RA Collige A., Vandenbergh I., Thiry M., Lambert C.A., Van Beeumen J.,
RA Li S.W., Prockop D.J., Lapiere C.M., Nussgens B.V.;
RT "Cloning and Characterization of ADAMTS-14, a Novel ADAMTS Displaying
RT High Homology with ADAMTS-2 and ADAMTS-3.";
RL J. Biol. Chem. 277:5756-5766(2002).
DR EMBL: AF366351; AAL79814.1; -.
SQ SEQUENCE 1159 AA; 127336 MW; A5B130149BF7F34 CRC64;

Query Match 30.2%; Score 616; DB 4; Length 1159;
Best Local Similarity 36.2%; Pred. No. 1.3e-52;
Matches 135; Conservative 60; Mismatches 156; Indels 22; Gaps 11;

QY 8 VETLVVADESVMKFGAD-LEHYLLTLTATAARLYRHPHSILNPINIVVVKVLLLRDRSG 66
Db 194 IEVLLVDDSVVRFHKGHEVQVYVLTLMNIVDEIYHDESLGVHINIALVRLIMVGYRQL 253
QY 67 PKV-TGNAALTILRNFCAMQ-KKLNVSKHPEYD-TAILTRDQLCGATTCDTLGMADVT 125
Db 254 SLIERGNFSRLEQVCRWAHSOORODPSAHHHDHVFLTRQDF-GPSGMO--GVAPVTG 310
QY 126 MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHD---NVKCEEVFGKLRANHMSPSL 182
Db 311 MCHPLRSCALNHEDGFSFAFVIAHETGHVLMGHEHDGQGCADFTSLGS-----VMAPLV 365
QY 183 IQIDRANPHSACSAALITDFLDLSHGDCLLDOPSKPI-SLPEDLPGASYTSLQCELAFG 241
Db 366 QAAHFRHWSRCSKLESLRPLS--YDCLDDPDPFAPWQPPELPFGINYSMDQCREFDFG 423
QY 242 VGSKPC---PYMQYCTKLWCTCKAKGMQVCTPRHPWADGTSCGEGKCLAGACVERHNL 298
Db 424 SGYQCLAFRTPECKQLWCS-HPDNPTFKTKGPPDLDGTECAPKWCFFKGHCITWKSPE 482
QY 299 NKHRVDGSAKWDYPCSRCTCGGVQVLARQCTNPTPANGKYCEGVVRVYRSCNLEPPC 358
Db 483 QTYGQDGGSSWTKFGSCRSOGGVRSSRCNNPSPAYGGRPCLGPMFEYQVCNSEEC 542
QY 359 PSSASGKSPREQ 371
Db 543 PGTY--EDFRAQ 553

RESULT 13
Q8TE59 PRELIMINARY; PRT; 1207 AA.
AC Q8TE59;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ADAMTS-19.

```

```

GN ADAMTS19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL: AJ311904; CAC84565.1; -.
SQ SEQUENCE 1207 AA; 134061 MW; AF36F6BF5886FDE2 CRC64;

Query Match 29.7%; Score 606; DB 4; Length 1207;
Best Local Similarity 36.8%; Pred. No. 1.4e-51;
Matches 142; Conservative 58; Mismatches 148; Indels 38; Gaps 15;

QY 8 VETLVVADESVMKFGAD-LEHYLLTLTATAARLYRHPHSILNPINIVVVKVLLLRDRSG 66
Db 327 IETVVADPAMVSYHGADAARRFTILTNMVFNLFOHKSGLGVQVNLKRVILIHETPPE 386
QY 67 PKVTGNAALTILRNFCAMQ-KKLNVSKHPEY---W-----DTAILTRDQLC--GAT 113
Db 387 LYIGHGHEKMLESPCKQKQHEEFKKNIDHLEMSTNGEDMTSVDAAILITRKDFCVHKDE 446
QY 114 TCDPLGMADVGTCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDN-VKVCCEVFGKL 172
Db 447 PCDTVGIAYLSGMCSEKRCIIAIDNGLNLAFTTAHEMGNHMGINDHNDHPSCAD----- 501
QY 173 RANHMSPSLIQIDRAN--PWSACSAALITDFLDLSHGDCLLD---QPSKPSILPELPG 227
Db 502 -GLHIMGEWIKGNLGDVSNRCSKEDLERLSKASNCCLLTNPQSNVSMVPSKLP 560
QY 228 ASYLSQOCELAFLGVGSKPCPYMQY--CTKLWCTGKAGQMVQCTPRHPWADGTSCGEGK 285
Db 561 MTTATDEQCOILFLGPLASFQCEQMHVICTGLWC--KVEGEKEGRTKLDPPMDGTDCLGK 618
QY 286 LCLKAGACVERHNLNKHRYDGSNAKWDYPCSRCTCGGVQVLARQCTNPTPANGKYCEG 345
Db 619 WCKAGECTSRTPSEH-LAGEWSLW---SPCSRCSAGISSRERKC--PGLDSEARDNG 672
QY 346 VRVYRSCNLEPPCSPSSASGKSPREQ 371
Db 673 PRKYRICENPPCPAGLPG--FRDQ 696

RESULT 14
Q95N24 PRELIMINARY; PRT; 192 AA.
AC Q95N24;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE AggreCanase-1 (Fragment).
DE AggreCanase-1 (Fragment).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Flannery C.R., Little C.B.;
RT "Expression and activity of equine aggrecanases.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF368321; AAK53425.1; -.
DR MEROPS: M12.221; -.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000130; Zn_Mtpeptidse.
DR PROSITE: PS00215; ADAM_MEPRO; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.

```

```
FT NON_TER 1 1
FT NON_TER 192 192
SQ SEQUENCE 192 AA; 20670 MW; 9013B0E19FCE8C56 CRC64;

Query Match 27.3%; Score 558; DB 6; Length 192;
Best Local Similarity 51.0%; Pred. No. 8.2e-48;
Matches 98; Conservative 26; Mismatches 64; Indels 4; Gaps 3;

QY 120 MADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGK-LRANHMM 178
DB 1 MADVGTCDPARSCAIVEDDGLSATAAHELGHVFNMLHDNSKPCVGLNGPSTSRHVM 60

QY 179 SPTLIQIDRANPWSACSAAIITDFLDSGHGDCLLDQPSKPISLPEDLPGASYTLSSQCEL 238
DB 61 APVMAHVDPEEPSPCSAREITDFLDNGYGHCLLDKPEAPLHPVTFPGKDYDADRQCQL 120

QY 239 AFGVGSFPCPYM-QYCTKLNCTGKAKGQVCQTRHFPWADGTSCGEGKCLCKAGACVERHN 297
DB 121 TFGPDSRHCFPLPPPCAAALWCSGHLNGHAMCQTKHSPWADGTFCGPAQACMGGRCLHMDQ 180

QY 298 LNKHRVD--GSW 307
DB 181 LQEFNIPQAGGW 192

RESULT 15
Q96L37 PRELIMINARY; PRT; 1427 AA.
AC Q96L37
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Von Willebrand factor-cleaving protease precursor.
GN ADAMTS13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RX PubMed=11557746;
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,
RA Fujikawa K.
RT "Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a
RT Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.";
RL J. Biol. Chem. 276:41059-41063(2001).
DR EMBL: AY055376; AAL17652.1; -.
DR MEROFS; M12.241; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tspl; 4.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPI; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR KW Protease; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 75 1427 VON WILLEBRAND FACTOR-CLEAVING PROTEASE.
SQ SEQUENCE 1427 AA; 153632 MW; E81BC3AABCA14442 CRC64;

Query Match 26.4%; Score 540; DB 4; Length 1427;
Best Local Similarity 32.1%; Pred. No. 7.1e-45;
Matches 127; Conservative 45; Mismatches 180; Indels 44; Gaps 8;

QY 7 YVETLVVADESMVKFHGADLEHLTLTLLATAARLYRHPSTILNFINIVVYKVLRLDRDSC 66
DB 81 HLELLVAVGPDVFOAHQEDTERVYLTNLNTGAEELRDRPSLGQAFRHLVKNVILTEPEGA 140

QY 67 PKVTGNAALTIRNFCAMQKLNKVSXKHPEYWDTAIFLTRQDL-CGATTCDTLGMADVGT 125
DB 141 PNITANLTSSLLSVCWSQINPEDDDTDPGHADLVLYITRFDLELDPGNRQVRGVTQLGG 200
```

```
QY 126 MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGKLRANHMMSPTLIQI 185
DB 201 ACSPTWSCSLITEDTGFDLGVTIAHEIGHSGFLEHDGAPGS-----CGGPSGHVMAS----- 251

QY 186 DRANP-----WSACSAAIITDFLDSGHGDCLLD-----QPSKPISLPEDLPGASYTLSSQOC 236
DB 252 DGAAPRAGLAWSPCSRRQLLSLLSAGRARCVNDPPRPQPGSAGHPDPAQAPGLIYSANEQC 311

QY 237 ELAFGVGSKPCPY-----MOYCTKLNCTGKAKGQVCQTRHFPWADGTSCGEGKCLCKAGAC 292
DB 312 RVAFGPKAVACTFAREHLDLMOQALSCHTDPLDQSSCSRLVPLLDGTGCGEKKWCSKGRG 371

QY 293 VERHNLNK-HRYDGSWAKWDYPGCSRTCGGGVQLARRQCTNPTPANGKCYCEGVRYVKYR 351
DB 372 RSLVELTPIAAVHGRSSWGPSPCSRSCGGGVVTRRQCNNPRPAPFGGRACVADLQAE 431

QY 352 SCNLEPCP-----SSASGKSF 367
DB 432 MCNTQACEKTQLEFMSQOCARTDQPLRSPGASG 467
```

Search completed: April 29, 2003, 17:17:45
Job time : 31.6233 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 11.9712 seconds
(without alignments)
1813.869 Million cell updates/sec

Title: US-10-009-332-1-copy_213_950

Perfect score: 4043
Sequence: 1 FVSIPIRYVETLVVADESMVK.....DQCNLRKPKQLDFCVLRPC 738

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2254	55.8	967	US-09-130-491-2	Sequence 2, Appli
2	1911.5	47.3	905	US-09-369-364A-9	Sequence 9, Appli
3	1900	47.0	608	US-09-130-491-13	Sequence 13, Appli
4	1761	43.6	837	US-09-122-126B-2	Sequence 2, Appli
5	1757.5	43.5	1882	US-09-369-364A-13	Sequence 13, Appli
6	1745	43.2	930	US-09-122-126B-15	Sequence 15, Appli
7	1730.5	42.8	874	US-09-369-364A-15	Sequence 15, Appli
8	1726	42.7	930	US-09-369-364A-2	Sequence 2, Appli
9	1607	39.7	551	US-09-130-491-16	Sequence 16, Appli
10	1273	31.5	481	US-09-130-491-8	Sequence 8, Appli
11	1264	31.3	518	US-09-369-364A-22	Sequence 22, Appli
12	1139.5	28.2	1224	US-09-930-872-4	Sequence 4, Appli
13	1084	26.8	997	US-09-369-364A-7	Sequence 7, Appli
14	1060	26.2	1211	US-09-491-522-5	Sequence 5, Appli
15	1033	25.6	1081	US-09-369-364A-17	Sequence 17, Appli
16	1023.5	25.3	1205	US-09-491-522-11	Sequence 11, Appli
17	925.5	22.9	859	US-09-369-364A-5	Sequence 5, Appli
18	846.5	16.0	245	US-09-369-364A-11	Sequence 11, Appli
19	550.5	13.6	525	US-09-369-364A-21	Sequence 21, Appli
20	421.5	10.4	566	US-09-491-522-7	Sequence 7, Appli
21	380.5	9.4	450	US-09-369-364A-19	Sequence 19, Appli
22	332	8.2	491	US-09-930-872-2	Sequence 2, Appli
23	268	6.6	751	US-08-836-443-3	Sequence 3, Appli
24	248.5	6.1	1170	US-08-313-288B-20	Sequence 20, Appli
25	240	5.9	441	US-08-985-526-3	Sequence 3, Appli
26	237	5.9	812	US-09-632-098-4	Sequence 4, Appli
27	227	5.6	802	US-09-632-098-2	Sequence 2, Appli

28	226.5	5.6	855	4	US-09-813-819-2	Sequence 2, Appli
29	226.5	5.6	855	4	US-09-920-048-2	Sequence 2, Appli
30	219	5.4	814	4	US-09-813-819-4	Sequence 4, Appli
31	219	5.4	814	4	US-09-920-048-4	Sequence 4, Appli
32	211.5	5.2	529	2	US-08-836-442-3	Sequence 3, Appli
33	200.5	5.0	788	2	US-08-918-914-4	Sequence 4, Appli
34	199	4.9	616	4	US-09-608-790-1	Sequence 1, Appli
35	199	4.9	1172	1	US-08-313-288B-19	Sequence 19, Appli
36	196.5	4.9	1239	2	US-08-937-931-2	Sequence 2, Appli
37	196.5	4.9	1239	4	US-09-285-502-2	Sequence 2, Appli
38	196.5	4.9	1239	4	US-09-709-126-2	Sequence 2, Appli
39	196.5	4.9	1239	4	US-09-871-385A-2	Sequence 2, Appli
40	193.5	4.8	239	5	PCT-US93-01652-1	Sequence 1, Appli
41	187.5	4.6	613	4	US-09-026-001A-10	Sequence 10, Appli
42	187.5	4.6	621	4	US-09-026-001A-6	Sequence 6, Appli
43	187.5	4.6	621	4	US-09-026-001A-18	Sequence 18, Appli
44	184	4.6	201	4	US-09-411-329C-1	Sequence 1, Appli
45	184	4.6	201	4	US-09-411-335-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-130-491-2

Query Match	55.8%	Score 2254;	DB 4;	Length 967;
Best Local Similarity	54.0%	Pred. No. 3.3e+185;		
Matches	403;	Conservative 131;	Mismatches 172;	Indels 40; Gaps 11;
QY	1	FVSIPIRYVETLVVADESMVKFHGADLEHYLTTLATAARLYRHPISILNPINIVVKKVLL	60	
DB	253	FVSSHRVYVETLVVADESMVKFHGADLEHYLTTLATAARLYRHPISILNPINIVVKKVLL	312	
QY	61	RDRSGPKVTGNAALTNRNCAWOKLANKVSDKHPEYWDTAITLFTRODLCGATTCDTLGM	120	
DB	313	HDEQKPEVTSNAALTNRNCAWOKLANKVSDKHPEYWDTAITLFTRODLCGATTCDTLGM	372	
QY	121	ADVTGMDCKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP	180	
DB	373	ADVTGMDCKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP	432	
QY	181	TLIQIDRANPNWASACNAITITDLSHGDCDCLLDQPSKISLPEDLPASYSYLSQOCELAF	240	
DB	433	MLSNLDHSPQSPCSAYMITITDLSHGDCDCLLDQPSKISLPEDLPASYSYLSQOCELAF	492	
QY	241	GVGSKPCP-YMOYCYKLMWCTGKAGQWVQTRHFPWADGTSCGEGKLCGACVVERINLN	299	
DB	493	GVGSKPCP-YMOYCYKLMWCTGKAGQWVQTRHFPWADGTSCGEGKLCGACVVERINLN	551	
QY	300	KH---RVDGSAKAWDYPGCSRCTCGGGVOLARRQCTNPTPANGKYGCVGVKVRYSNLE	356	
DB	552	KHFDTPFHGSMGWMGPGWDCSRCTCGGGVOLARRQCTNPTPANGKYGCVGVKVRYSNLE	611	

QY	357	PCSSASGKFRBEOCEAFNGYNHSTNRLTLAVAVPKYSGVSPDKCKLICRANGTYF	416
Db	612	DCPDN-NGKTFREOCEAHNEFSAKSGFSGPAVEWPKYAGVSPDKRCKLICOAKIGYF	670
QY	417	YVLAPKVVDTCLGSPDSTSVCGKCIKAGCDNLGSKKRFKDCGCGGDNKSKKVTGL	476
Db	671	FVLPKVVDTGSPDSTSVCGQCVKAGCDRIIDSKKFKDCGCGGNGSTCKKISGS	730
QY	477	FTKPMHGFVVAIPAGASSIDIRQYKGLIGDDNYLALKNSQGYLNGHVFVSAVER	536
Db	731	VTSAPKGYHDIITPTGATNIEVKQRNQRNRNGSFLAKAADGYIILNGDYTLSTLEQ	790
QY	537	DLVVKGSLLRYSGTGTAVESLQASRPILPLTVSVGKWTTPRVYSFYLPEKPREDK	596
Db	791	DIMYKGVLYRSGSSALERISFSPLKEPLTTQVLTGVNALRPKITYTFVKKK----	846
QY	597	SSHPKDPGRPSVLHNSVLSNOVEQDDRPAPRWAGSMGPGCSASGSGLOKRAYDCRG	656
Db	847	ES-----FNAITPFS-----AWTEEWGECSKCELGWQRRLVECRD	883
QY	657	SAGQRTVPACDAH--RPVETQAGE-PCPTWELSAMSPCKSGRGFRQRSLKCVGHGG	713
Db	884	INGQ---PASECAKEVPASTRPCADHPQWQLGEWSSCKTGKGYKRSKLCLSHDG	940
QY	714	RLARDQCNLHRKPQE-LDFCVLRPC	738
Db	941	GVLSHESCDPLKKPKHFDICTMAEC	966
RESULT 2			
US-09-369-364A-9			
; Sequence 9, Application US/09369364A			
; Patent No. 6391610			
; GENERAL INFORMATION:			
; APPLICANT: Apte, Suneel			
; APPLICANT: Hurskainen, Tiina L.			
; APPLICANT: Hirohata, Satoshi			
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases			
; FILE REFERENCE: 26473/4007/10-30-00			
; CURRENT APPLICATION NUMBER: US/09/369,364A			
; CURRENT FILING DATE: 1999-08-06			
; NUMBER OF SEQ ID NOS: 31			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 9			
; LENGTH: 905			
; TYPE: PRT			
; ORGANISM: Mus musculus ADAMTS-8			
US-09-369-364A-9			
Query Match 47.3%; Score 1911.5; DB 4; Length 905;			
Best Local Similarity 50.6%; Pred. No. 8.8e-156;			
Matches 352; Conservative 127; Mismatches 185; Indels 31; Gaps 13;			
QY	1	FVSIPIRYVETLVADESVMVKFHGADLEHYLTLLATARLYRHPHSILNPINIVVKVLLL	60
Db	229	FYSEARFVETLVAASMAAFYGTDLQHILTVMSMAARIYKHPISRNVLNVVKVLLV	288
QY	61	RDRDSGPKVTGNAALTNRNCAWKLNKVSOKHPEYWDTAILFTFRQDLGCG-ATTCDTFLG	119
Db	289	EXERGPEVSDNGGLTLNRNCSQWRFNKPSDRHPEHYDTAILFTFRQFCGKGEOCDTLG	348
QY	120	MADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNPHDNKVKCEEVFKGLRANHMS	179
Db	349	MADVGTICDPKSCSVIKDEGLQAAYTLAHELGHVLSLPHDDSKPCVRLFGPMGYHMA	408
QY	180	PTLIQIDRANPHSACSAIITDFLDSGHGCDLLDQPSKPSIPEDLPGAS--YTLSSQCE	237
Db	409	PFYIHNKTLNTPSPCSAVYLTLELDDGHGCDLLDAPTSLVLPFTGLPGHSLYELDQOCK	468
QY	238	LAFGVSKPCP---YMQYCTKLWCTGKAGQVMVQTRH--FPWADGTSCGSKLCKGAC	292
Db	469	QIFGPDPRHCPNTSVEDICVLQCARHRSDEPICKNGKSLWLWADGTPCGPGHLCLDGSC	528
US-09-130-491-13			
; Sequence 13, Application US/09130491			
; Patent No. 6416974			
; GENERAL INFORMATION:			
; APPLICANT: Holtzman, Douglas A.			
; APPLICANT: Goodearl, Andrew D.J.			
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83			
; FILE REFERENCE: 09404/041001			
; CURRENT APPLICATION NUMBER: US/09/130,491			
; EARLIER FILING DATE: 1998-08-07			
; EARLIER FILING DATE: 1997-09-05			
; EARLIER APPLICATION NUMBER: US 60/054,961			
; EARLIER FILING DATE: 1997-08-06			
; NUMBER OF SEQ ID NOS: 16			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 13			
; LENGTH: 608			
; TYPE: PRT			
; ORGANISM: Mus musculus			
US-09-130-491-13			
Query Match 47.0%; Score 1900; DB 4; Length 608;			
Best Local Similarity 49.3%; Pred. No. 4.9e-155;			
Matches 339; Conservative 114; Mismatches 146; Indels 88; Gaps 8;			
QY	57	VLLLRDRSGPKVTGNAALTNRNCAWKLNKVSOKHPEYWDTAILFTRODLGGATTC	116
Db	4	ILTVHDKQKPEVTSNAALTNRNCAWKLNKVSOKHPEYWDTAILFTRODLGGATTC	63
QY	117	TLGMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNPHDNKVKCEEVFKGLRANH	176
Db	64	TLGMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNPHDNKVKCEEVFKGLRANH	123
QY	177	MMSPTLIQIDRANPHSACSAIITDFLDSGHGCDLLDQPSKPSIPEDLPGASYSLSQOC	236
Db	124	MMSPTLIQIDRANPHSACSAIITDFLDSGHGCDLLDQPSKPSIPEDLPGASYSLSQOC	183
QY	237	ELAGVSKPCP-YMQYCTKLWCTGKAGQVMVQTRH--FPWADGTSCGSKLCKGACVER	295
Db	184	QTFEGEDSKHCPDAASTCTSLWCTGSGVILVQCTKHFHWADGTSCGEGKWCINGKVCN	243


```

Db 243 GLAELCTICDPYRSCSISDSGLSTAFTHAELGHVFNPHDDSNCKKEE-GVKSPQHV 301
QY 179 SPTLIQIDRANPWSACSAIIITDFLDSHGDCLLDOP-SKPSISLPELPGASYTISQOCE 237
Db 302 APTLNFTYTPNMMWMSKSKRYITFEEDLTGYGECCLNEPASRTYPLPSQLGLLYNKNQCE 361
QY 238 LAFVGSKPCPYMOYCTKLWCT--GKAGOMVCOTRIFFPWADGTSCEGKLCCLKAGAVE 294
Db 362 LIFPGSQVCYPMQCRRLLWCNNVDGAHGK---CKTQHTPWADGTECEPBGHCKRGFCVP 418
QY 295 RHNLKHRVDGSAKWDPYPCSRCTCGGQVQLARRQCTNPTPANGKCYCEGVYKYSRSCN 354
Db 419 K-EMEGPAIDGSGWGSHEFTGSRCTCGGKIKTAIRECNRPKNGKCYCVRMKFKSCN 477
QY 355 LEPSPSSAGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSVSPRDKCKLICRANGTG 414
Db 478 TEFCKMQK--RDFREOCEAFNGKHNGLNLLPSVRWFPKYSIGILMKDRCCLKFCRVAGNT 535
QY 415 YFVYLPKVVDTGLCSPTSTSVCGKCIKAGCDGNLGSKKRFRDKCGVCGGDNKSKKVT 474
Db 536 AYQLRDRVIDTGPCQDNDICVOGLCRQAGCDHILNSKVRKDKCGICGGDNSSCKTVA 595
QY 475 GLFTKPMHYNFVAIPAGASSIDTRQRYKGLIGDDNYLALKNSQGYLLNGHFVVSAY 534
Db 596 GTFNTVHYGYNVVRIPAGATSIDYRQHSFSGKSEDDNYLALNSKGBEFLNGDPVWSMS 655
QY 535 ERDLVVKGLLYRTGTGTAVESLQASRPITLPLETVLSVGMKTPPRYRYSFLPKPRE 594
Db 656 KREVRGSAVIEYSGDNNVCELNCTDRIEELLQVLSVGLYNPDYRYSFNPIEDKP 715
QY 595 DK---SSH-----PKDPRGSPVHLNHSVLSLSNQ-----VEQPDDRPPA----- 629
Db 716 QQFYWNHSGPMQACSKPCOGERRPKLVCTRESQDQTVSDQCDRLPQPGPYTEACGTDCD 775
QY 630 -RWVAGSWGPCSAACGSLQKRAVDRC-----GSAGQRTVPACDAARHPVETQACGPPC 682
Db 776 LRHWVASKSECAQCGGLGYRTLDIHCAKYSRMDGKTERKVDVDFSCSQPRPSNOBKSGEC 835
QY 683 PT--WELSAWSPCSKCGRGFORRLKCV 709
Db 836 STGGWRYSAWTECSRSCDGGTHRRRAICV 864

RESULT 8
US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

Query Match 42.7%; Score 1726; DB 4; Length 930;
Best Local Similarity 47.4%; Pred. No. 8.3e-140;
Matches 327; Conservative 105; Mismatches 218; Indels 40; Gaps 9;

QY 2 VSPRYVETLVVADSMVKFAGLEHYLLTLATAARLYRHPISLNPINTVWVKVLLLR 61
Db 263 ISRARQVELLVADSMVARMYGRGLQHYLLTWASIANLYSHASINHLAVKVVVLT 322
QY 62 DRDSGPKYTGNAULTLRNFCARQKLNKYSKHPYEDWTAILFTQRDLGCGATTCDTLGMA 121

```

```

Db 323 DKDTSLEYSKNAATYLANFCKWQHNLQDGDHHEHYDAALLFTREDLCGHSCDTLGMA 392
QY 122 DVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNYKVEEYFGKLRANHMMSPT 181
Db 383 DVGTCSPERSCAVIEDDGLHAFTVAHEIGHLLGLSHDDSKFCEENFGTTEDKRLMSSI 442
QY 182 LIOIDRANPWSACSAIIITDFLDSHGDCLLDOPSKPSISLPELPGASYTISQOCELAPG 241
Db 443 LTSIDASKPSKCYSAITIFELDDHGNCLLDLPKRKILGPPELPGQTYDATOCNLTFG 502
QY 242 VGSKPYMYCTYKLTWCTGKAGOMVCOTRIFFPWADGTSCEGKLCCLKAGAVE-----H 296
Db 503 PEYSVCPGMDYCARLCAVAVRQGMVCLTKKLPAVEGTGCGKGVCLQKGCVDKTKKYY 562
QY 297 NLNKHVRVDGSAKWDPYPCSRCTCGGQVQLARRQCTNPTPANGKCYCEGVYKYSRSCNLE 356
Db 563 STSSH--GNWGSMPWQCSRSCGGVQFAYRHCNNPAPRNSGRYCTGKRATYRSCSVT 619
QY 357 PCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSVSPRDKCKLICRANGTGTF 416
Db 620 PCP--PMGKSFRRHQCEAKNGYOSDAKGVKTFVWVPKYAGVLPADVCKLTCRAKGTGY 677
QY 417 YVLAPKVVDTGLCSPTSTSVCGKCIKAGCDGNLGSKKRFRDKCGVCGGDNKSKKVTGL 476
Db 678 WFSPKVTDTGTECPYNSVCRGRCVTRTCDGLIGSKLOYDKGCVCGGDNSSCTKIIGT 737
QY 477 FTKPMHYNFVAIPAGASSIDTRQRYKGLIGDDNYLALKNSQGYLLNGHFVVSAYER 536
Db 738 FNKSKGYTDVRIPEGATHIKVRQFAKQDTRFPAALALKKTKTEYLINGKMTSET 797
QY 537 DLVVKGLLYRTGTGTAVESL-----QASRPILLEPLTVLSVGMKTPPRYRYSFLPKPE 591
Db 798 IIDINGTVMYNSGSHRDDFLHGMYSATKEI---LIVQILATDPTKALGVRYSFVFPKK 854
QY 592 PREDKS---SRPKDPRGSPVHLNHSVLSLSNQVEQPDPRPARVAGSWGPCSAACGSLQ 648
Db 855 TTQKVNVSIVSHGSKVGP---HSTQL-----QWVTGPWLACSTCDTGTWH 896
QY 649 KRAVDCRGSAQRTVPACDAARHPVETQAC 678
Db 897 TRTVQCO-DGNRKLAKGCLLSQRPSAFKQC 925

RESULT 9
US-09-130-491-16
; Sequence 16, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Gooden, Andrew D. J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-130-491-16

Query Match 39.7%; Score 1607; DB 4; Length 551;
Best Local Similarity 48.5%; Pred. No. 6.8e-130;
Matches 282; Conservative 108; Mismatches 152; Indels 40; Gaps 10;

QY 165 CEEVFGKLRANHMSPFLIOIDRANPWSACSAIIITDFLDSHGDCLLDOPSKPSISLPE 224

```

```
Db 1 CASLNGVSGDHLMSLSDHSQSPSCSAITVMTSFLDNGHGECMLDKPQNKLPSD 60
Qy 225 LPGAASYLSOOCELARGVSGKPCP-YMQYCTKLWCTGKAKGOMVQCTREHPWADGTSCE 283
Db 61 LPTGLDANQCOQFTGEESKHCPCDAASTCTTLCWCTGTSGLLVCTQKTFPPWADGTSCE 120
Qy 284 GKLCAGACVERHNLNKH---RYDGSWAKWDYGPSCSRTCGGGVQLARRQCTNPTPANG 340
Db 121 GKVCVSGKCNKTDK-KHFATPVHGSWGPWPGDCSRTCGGGVQVQTMRECDNPVPKNG 179
Qy 341 KYCEGVRYKRSNLEPCSPSSASGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSGVSP 400
Db 180 KYCEGRVRYRSCNIEDCPDN-NGKTFREEQCEAHNEFSKASFGNEPTVEWTPKYAGVSP 238
Qy 401 RDKCKLICRANGTGYFYVLAPKVVDGTLGSPDSTSCVCGKCIKAGCDGNLGSKKRFRDC 460
Db 239 KDRCKLTCEAKGIGYFVLPQKVVDGTPCSPDSTSCVCGQCVKACDRDIIDSKKFKDC 298
Qy 461 GVCGGDNKSKCKVTGLTKPMHGYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQ 520
Db 299 GVCGGNGSTCKKMSGIIVSTRPGYHDVITIPAGATNIEVKHRNORGSRNNGSFLAIRAD 358
Qy 521 GKYLNGHFVVAVERDLVVGSLRYSGTAVESIOASRPILPILTVVEVLSVGRKWTTP 580
Db 359 GTYLNGFTLSLEQDLTKVLYRYSASAALERIRSFSLKEPLTIQVLVGHALRP 418
Qy 581 RVRYSFYLPKPREDKSSHPKDRPGPSVLHNSVLSLNOVEQPDPRPPARWAGSWGPCS 640
Db 419 KIRFTYFMKKKTES-----FNAIPTES-----EWVIEEWGEC 451
Qy 641 ASCSGSLQKRAVDCRGSGAGORTVPACDAH--RPVETOACGE-PCPTWELSAWSPCKSC 697
Db 452 KTCGSGQWRVVOCRDINGH---PASECAKEVKPASTRPCADLPCPHWQVGDWSPGSKTC 508
Qy 698 GRGQRESLCKVGHGGHLLARDONLHRKPQE-LDFCVLRPC 738
Db 509 GKGYKRTKLKCVSHGGVLSNESCDPLKPKHYIDFCTLTCQ 550

RESULT 10
US-09-130-491-8
; Sequence 8, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; EARLIER FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-130-491-8

Query Match 31.5%; Score 1273; DB 4; Length 481;
Best Local Similarity 47.0%; Pred. No. 3e-101;
Matches 236; Conservative 87; Mismatches 141; Indels 38; Gaps 12;

Qy 190 PWSACSAIITDFDLSHGDCLLDQPSKPISLPDLPGASYLSQCELAFVGSKPCP- 248
Db 7 PWSPCSAVYLTFLDLDGHDGDCLLD-----DGHSTLYELDQCKQIFGPDFRCPN 56
Qy 249 --YMQYCTKLWCTGKAKGOMVQCTRH--FPWADGTSCEGKLCUKGACVERHNLN--KHR 302
Db 57 TSVEDICVQLWCRHRSDEPICHTKNLSLLWADGTPCGPHLCLDGLDSCVLRVEENPKAV 116
```

```
Qy 303 VDGSWAKWDYGPSCSRTCGGGVQLARRQCTNPTPANGKYCEGVRYKRSNLEPCPSA 362
Db 117 VDGWGPWGPWGQCSRTCGGGIQFSNRECDNPAPQNGRFRCLGERVKYQSKTECP--P 174
Qy 363 SGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSGVSPDKCKLICRANGTGYFYVLAP 422
Db 175 NGKSFREOCCQKYNAYNH-TLDNGNFTLQWVPKYSGVSPDRCKLFCRARGSEKVFETK 233
Qy 423 VVDGTLGSPDSTSCVCGKCIKAGCDGNLGSKKRFRDCGVCGGDNKSKCKVTGLTKPMH 482
Db 234 VIDGTLGCPDLAICVRQCVKAGCDHVNNSPKLDCGVCGGKGTACRKVSGSFTPEY 293
Qy 483 GYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQKYLNLNGHFVVAVERDLVVG 542
Db 294 GYNDVTITIPAGATNIDVQRSHPGVQNDGSLALKTANGQYLLNGLNLAISAIEDILMK 353
Qy 543 SLLRYSYSGTAVESIOASRPILPILTVVEVLSV-GKMTPPRVRYSFYLPKPREDKSSHPK 601
Db 354 TILKYSGSMATLERIQSFQALPELTVQLLVTSGEVFPKPKYTFVFPND--TDFNVQSS 411
Qy 602 DPRGPSVLHNSVLSLNOVEQPDPRPPARWAGSWGPCSASCSGLQKRAVDCRGSGAGOR 661
Db 412 KERAST--NIQSL-----PYAEWVLGDWSECPSTCGGQWRRTVECDRPPSGQA 458
Qy 662 TVPACDAAHRPVETQACG-EPC 682
Db 459 S-DTCDCAALPKEDAKPCGSGQPC 479

RESULT 11
US-09-369-364A-22
; Sequence 22, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-5
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (99)
; OTHER INFORMATION: Xaa = M
US-09-369-364A-22

Query Match 31.3%; Score 1264; DB 4; Length 518;
Best Local Similarity 45.4%; Pred. No. 2e-100;
Matches 244; Conservative 73; Mismatches 185; Indels 36; Gaps 10;
```

```
Qy 152 GHVFNNPHDNVYCEEVFGKLRANHMMSPTLIQIDRANPWSACSAIITDFDLSHGDC 211
Db 1 GHLGLSHDDSKFCEETFGSTEDKRLMSSILTSIDASKPWSKTSATITEFLDDHGNC 60
Qy 212 LDQPSKPISLPDLPGASYLSQCELAFVGSKPCPYMYCTKLWCTGKAKGOMVQCTR 271
Db 61 LDLPRKQILGPPELPQGTVDATQCCNLTFGPDYSVPCGXDXVCARLCAVAVROGVCLTK 120
Qy 272 HEPWADGTSCEGKLCUKGACVER-----HNLNKRHVDGWSWAKWDPYGPCSRTCGGGVQL 326
Db 121 KLPAVEGTGCGGRICLCQCKVDKTKKKYYSSSH---GNWGSWGSWGCSCSGGGVQF 177
Qy 327 ARRQCTNPTPANGKYCEGVRYKRSNLEPCSPSSASGKSFREOCEAFNGYNHSTNRL 386
Db 178 AYHCNPNAPRNGRYCTCKRAIYHSCSLMPCP--PNGKSFREOCEAKNGYQSDAKGVK 235
```

Db	637	SQKCPRDS--VDPRAAQCAEHNRRRFRGRHYK-----WKP-YTQVEDQDLCKLYCI	684
Qy	410	ANGTGYFVLAPKVDGTLCSPDSTSVCVQGRKICAKGCDNIGSKRRFDKGVGGDNKS	469
Db	685	AEQDFDFEFLSNKVKDGTPCSDERSNVICDGTICERVGCNDVLGSDAVEDVGCVCNGNSA	744
Qy	470	CKKVTLGFTKPMH--GYNFWAIPACASSIDIRQRCYKGLIGDDNYLALKNSQGYKLLNG	527
Db	745	CTTHRGLYTKHHHTNOYHYHMTIPSGARSIRYEMNV-----STSYISVRNALRRYYLNG	799
Qy	528	HFVVSAYERDLVKGSLRLRGSGTGAVERLSIQASRPTLEPLTVEVLVSQVKMTPPRVKSYFY	587
Db	800	HWTVDWPCR-YKFSGTTFDYRYSNEPENLIATGPTNETLIVELLFOGR--NPGVAWEYS	856
Qy	588	LPKEPREDKSSHKPDGRPGS-----VLHNSVLSLNSQVE-----QP	623
Db	857	MPR-----LGTEKQPPAQPSYTWAIVRSECSVSCGGQMTVEGCRYDLKLFQVNMSCFNP	911
Qy	624	DDRP-----PARVAGSWGPCASGSGLOKRAVDC--RGSAGORTYPA--CD	667
Db	912	KTRPVTLGPKCVKSYACPPPSWSVGNWSACSTRTCGGGAQSRPVQCTRHRVYDSEPVASLCP	971
Qy	668	AAHRPVETQAC-GEPC-PTWELSAWSPCKSGKGRGFRQRSLKCVGHG---GRLIARDQC	721
Db	972	QP-APSSRQACNSQSPPAWSAGPAWCSHTCTCGKWRKRAVACKTNP SARAQILLPDAVC	1030
Qy	722	NLHRKKPQELDFCVLRPC	738
Db	1031	TSEPKPRMHEACILLQRC	1047

```

RESULT 13
US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7

```

QY	8	VETLVVADESXWKFHG-ADLEHYLLTLATAARLYRHPHSILNPINTVVKVVLRLDRDSG	66
Db	292	VETLVVYDKMMQNGHENITTVLTILNMVSALPKDGTIGGNIAIVGLILLDEDEQBG	351
QY	67	PKYTGNAALTLRNFCWAQKKL-NKVSDDKIPEYWDTAIFLTRQDLG--ATTCDTIGMADV	123
Db	352	LVISHADHPTLSSFCQWQSLMGKDTRH---DHAILLTGLDICSWNKEPCDTLGFAP	407
QY	124	GTWCDPKRCSVTEDDGLSFAHTAEHGLHVENKPHDNV-KYCEEVFGKLRANHMSP	182
Db	408	SGMCSYRSCTINEDTGLGLAFTIAHSEGNHGMTHDGEWNCKKSEG-----NTMSPT	462
QY	183	IQIDRANPWSACAAITDFDLSGHGDCLLDOPSKPI--SLPEDLPGASYTLSSOCELA	239
Db	463	AGRNFGVFSWSPCSRQLYHLHFLSTAQAICLADQP-KPVKEYKYPEKUPGELYDANTCKWQ	521
QY	240	FGVSGSKPCPY---MOYCTKLWC--TGKAKGQWVCOTRHPFPWADGTCGSGEKLCLXGACVE	294
Db	522	FGEKAKLMLDFKKDICKALWCHRIGRK-----CETKFPMAEGTICGHDHMCRCGGQCVK	576
QY	295	RHNLNKHVRDGSWAKWDYPGCSRTCGGQVQLARRQCTNPTPANGSKYCEGYRVKYRSCN	354
Db	577	YGDEGPKPTHGHSWDSSWSPCSRCTCGGVSHRSRLCTNPKPSHGKGFCEGSTRILKLCN	636
QY	355	LEPCPSSAGSKSPREQC-----EAPNGYNHSTNRLTLAVAWVPKYSVSPRDKCKLICR	409

Qy	2	VSIPRYVETLVVADESVMVFHG--ADLEHYLLTLLATAARLYRHPISILNPINIVVVKVLLL	60
Db	238	VSKEKWCETLVVADAKWVEYHGQPVESYVLTIMNVAGLFHPDSIGNPIHITIVRLVLL	297
Qy	61	RDRSGPKVTGNAALTLRNFCAWQKLNKVSQKHPYWDTAITLFRQDLGCATT--CDTL	118
Db	298	EDEEEDLKITHHADNTLLKSFCKWOKSINKMGDAHPLHHDTAILLTRKDLCAAMNRPCETL	357
Qy	119	GMADVTCMDPKBSCSVIEDDGLPSAFTTAHELGHVENKPHD--NVKVCSEEVFGKLRAHIM	177
Db	358	GLSHVAGMCPHRSCSINEDTGPLAFTVAHELGHSHFGIHDGSGNDCPEV-GK--RPFI	414
Qy	178	MSPTLIQIDRANPWSACSAAITDFLDSGHGBCLLDQPSKP--ISLPEDLPGASTYTLQQC	236
Db	415	MSPOLLYDAAPLWRSRCSQYITRFIDRWGVLGCLDDPPAKDIDPSPYPGVLVDVSHQC	474
Qy	237	ELAFVGSGKPCPYW--QVCTKLWCTGAKGMQVCQTRHFPFWADGTSCGSGKLCCLGACVER	295


```

; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-10
; IS-09-369-364A-17

```

Query Match	25.6%;	Score 1033;	DB 4;	Length 1081;
Best Local Similarity	31.8%;	Pred. No. 4.2e-80;		
Matches 272;	Conservative 104;	Mismatches 319;	Indels 160;	Gaps 28;
QY	2	VSIPRYVETLVVADESVMVKFHG-ADLEHYLLTLTAAARLYRHPHSILNPINIVVVKVLL 60		
DB	213	VSRERYETMDVADKMVAYHGRDVEQYVLAINIVAKLFQDSLSGTSVNLVTRLILL 272		
QY	61	RDRSGPKVGNAAALTIRNFCAWQKL-----NKYSDKHPEYDFAILFTRODLC--G 111		
DB	273	TEDOPTLEITHHAGSLDSFCKWQKSVNHSGHGNAPENGVAHNDTAVLTRYDICIYK 332		
QY	112	ATTCDTLGMADVGMCDPKRSCVIEDDGLPSAETTAHELGHVFMHDKVYCEEVFG- 170		
DB	333	NKPGCTGLAR-WAECVSAREAAASMTLAATSVHHCHTEIGHTGMHHDGVGNSCGARGQ 391		
QY	171	---KLRANHMSPFLIIQIDRANP--WSACSAAIITDFLDSGHGDCLLDQPSK-PISLPED 224		
DB	392	DPAKLMAAHTMKT-----NPFWSNCRDYITISFLDSGLGLCLNRRPQDFVYPTV 444		
QY	225	LPGASYILSQOCELAFGVGSKPCPYMOYCYTKLWCTKAKGOMVCQTRHPWADGTSCEG 284		
DB	445	APGOAYDADGCRFQHGKYGSKQYKGYEVCSELWCLSKSNR---CITNSIPAAEGTLCQTH 501		
QY	285	KL----CLKACACVERHNLKRVDSWAKMDPYGPCSRTCGGVQGLARRCQTNTPANGG 340		
DB	502	TIDKGWCYKRCVP-FGSRPEGVGDAGPWTWGDGCSRTCGGVGSSSRHCDSPRTIGG 560		
QY	341	KYCEGVVYKVRSCNLEPCPSSASGKSPREOCEAFNG-----YNHSTNRLTLAVAWVP 393		
DB	561	KYCLGERRRHRSCNTDCCPGS--QDREVOCAEFDSIPFRGKYKKTVR----- 609		
QY	394	KYSGVSPRDKCLICRANGTYEYVLAPKVVDGTLCPDSTSCVQCGKCIKAGCDGNLGS 453		
DB	610	-GGVK---ACSLTSLAEGNFYTERAAAVDGPTRPDVTIDCVSGECKHVGCDDVLGS 665		
QY	454	KKRFDCVCGGDNKSKCKVTGLFT--KPMHGFYFVVAIPAGASSIDIQRGYKGLIGDD 511		
DB	666	DLREDKRVCGGDSACETIEGVFSPASPGAGYEDVVMIPKGVSHFIQD-----LNL 720		
QY	512	NYLALKNSQKYLNLGHFVVSAYERDLVVGSLRLRYSGTGTAVESLQASRIEPLFVEV 571		
DB	721	SHLAKGDQSLLEGLPGTGPQPHR-LPTAGTTFLQRGDPDQVQSLAALGINASLVMV 779		
QY	572	LSVGKMTPPRVRYSFYLP----- 589		
DB	780	LARTEL--PALRYFNFNAPIARDSLPPYSWHYAPWTKCSAOCAGGSQVQAVECRNQLDSSA 837		
QY	590	-----KEPREDK--SSHPRDP-----RGPSVLHNSVLSLSNQ 619		
DB	838	VAPHYCSAHSKLPKRQACRNTPECPDPVMVGNNSLCSRSRCDAGVRSSTVVCQRVSAEE 897		
QY	620	VEQ-----PDDRPPA-----RWVAGSVCPSASCGLQKRAVDCRSAGORTV 663		
DB	898	KALDSDACPOPRPVLVEACHGPTCPDPWATLDWSECTPCGPGGLRHRVWLCKSADQRSTL 957		
QY	664	PA--CDAAHPRVETQAG-EPCP--TWELSAWSPCKSCGRCGRFORSLKCVGHGGRLLAR 718		

GenCore version 5.1.4_p5_4578
 Copyright (c) 1993 - 2003 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: April 29, 2003, 17:12:21 ; Search time 16.4049 Seconds
 (without alignments)
 3604.758 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_950
 Perfect score: 4043
 Sequence: 1 FVSIPRYVELLVVADESWK.....DQNLHKKPQELDFCVLRPC 738

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4043	100.0	950	10	US-09-965-631-4
2	3136.5	77.6	823	9	US-10-163-316-2
3	2260	55.9	727	9	US-10-097-597-1
4	2260	55.9	727	9	US-10-097-580-1
5	2260	55.9	727	10	US-09-445-023A-1
6	2259	55.9	727	9	US-10-097-597-12
7	2259	55.9	727	9	US-10-097-580-12
8	2259	55.9	727	10	US-09-445-023A-12
9	2257	55.8	950	10	US-09-321-987B-4
10	2257	55.8	968	9	US-10-163-316-7
11	2254	55.8	967	12	US-10-105-929-2
12	1911.5	47.3	905	10	US-09-918-171A-9
13	1901	47.0	608	10	US-09-803-589-2
14	1900	47.0	608	10	US-09-803-589-8
15	1900	47.0	608	12	US-10-105-929-13
16	1779	44.0	1829	10	US-09-972-467-2
17	1760	43.5	837	9	US-10-174-590-352
18	1760	43.5	837	9	US-10-176-758-352
19	1760	43.5	837	9	US-10-175-737-352

20	1760	43.5	837	9	US-10-173-706-352	Sequence 352, App
21	1760	43.5	837	9	US-10-175-738-352	Sequence 352, App
22	1760	43.5	837	9	US-10-175-752-352	Sequence 352, App
23	1760	43.5	837	9	US-10-176-482-352	Sequence 352, App
24	1760	43.5	837	9	US-10-176-757-352	Sequence 352, App
25	1760	43.5	837	9	US-10-176-913-352	Sequence 352, App
26	1760	43.5	837	9	US-10-180-552-352	Sequence 352, App
27	1760	43.5	837	9	US-10-180-557-352	Sequence 352, App
28	1760	43.5	837	9	US-10-173-700-352	Sequence 352, App
29	1760	43.5	837	9	US-10-174-572-352	Sequence 352, App
30	1760	43.5	837	9	US-10-174-579-352	Sequence 352, App
31	1760	43.5	837	9	US-10-174-582-352	Sequence 352, App
32	1760	43.5	837	9	US-10-174-588-352	Sequence 352, App
33	1760	43.5	837	9	US-10-175-739-352	Sequence 352, App
34	1760	43.5	837	9	US-10-175-740-352	Sequence 352, App
35	1760	43.5	837	9	US-10-175-743-352	Sequence 352, App
36	1760	43.5	837	9	US-10-176-488-352	Sequence 352, App
37	1760	43.5	837	9	US-10-176-492-352	Sequence 352, App
38	1760	43.5	837	9	US-10-176-747-352	Sequence 352, App
39	1760	43.5	837	9	US-10-176-750-352	Sequence 352, App
40	1760	43.5	837	9	US-10-176-985-352	Sequence 352, App
41	1760	43.5	837	9	US-10-176-987-352	Sequence 352, App
42	1760	43.5	837	9	US-10-176-991-352	Sequence 352, App
43	1760	43.5	837	9	US-10-176-992-352	Sequence 352, App
44	1760	43.5	837	9	US-10-176-993-352	Sequence 352, App
45	1760	43.5	837	9	US-10-184-658-352	Sequence 352, App

ALIGNMENTS

RESULT 1
 US-09-965-631-4
 ; Sequence 4, Application US/09965631
 ; Patent No. US20020115842A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Friddle, Carl Johan
 ; APPLICANT: Hilbun, Erin
 ; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
 ; FILE REFERENCE: LEX-0241-USA
 ; CURRENT APPLICATION NUMBER: US/09/965,631
 ; CURRENT FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: US 60/236,689
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; TYPE: PRT
 ; LENGTH: 950
 ; ORGANISM: homo sapiens
 US-09-965-631-4

Query Match 100.0%; Score 4043; DB 10; Length 950;
 Best Local Similarity 100.0%; Pred. No. 2.4e-287;
 Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	FVSIPRYVELLVVADESWKFGADLEHYLLTLATAARLYRHPSILNPINIVVKVLLL	60
DB	213	FVSIPRYVELLVVADESWKFGADLEHYLLTLATAARLYRHPSILNPINIVVKVLLL	272
QY	61	RDRDSGPKVTGNAALTNRNCAWKKLNKVSQKHPEYWDTAILFTRODLGGATTCDTLGM	120
DB	273	RDRDSGPKVTGNAALTNRNCAWKKLNKVSQKHPEYWDTAILFTRODLGGATTCDTLGM	332
QY	121	ADVGTMCDPKSCSVIEDDGLPSAFTTAHEIGHVFNPHDNVKNVCEEVFGKLRANHMSP	180
DB	333	ADVGTMCDPKSCSVIEDDGLPSAFTTAHEIGHVFNPHDNVKNVCEEVFGKLRANHMSP	392
QY	181	TLIQIDRANPNASAAITITDLSHGDCILDDPSKPISELPDIPGASVTLSSQCEELAF	240
DB	393	TLIQIDRANPNASAAITITDLSHGDCILDDPSKPISELPDIPGASVTLSSQCEELAF	452
QY	241	GVGSKPCPYMOYCTKLWCTGKAKGMVQCQTRHFFWADTSCGEGKLCCLKGACVERHNLN	300

Db 453 GVSGKPCPYMQYCTKLCTGAKGQWVCOTRHFPAWADGTCGSGKCLKGACVERHNLNK 512
QY 301 HRVDGSAWAKDYPGCSRTCGGQVQLARRQCTNPTPANGGKYCEGVVRYRSCNLEPCPS 360
Db 513 HRVDGSAWAKDYPGCSRTCGGQVQLARRQCTNPTPANGGKYCEGVVRYRSCNLEPCPS 572
QY 361 SASGKSFEQCEAFNGYNHSTNRLTLAVAVPKYSGVSPDRCKLICRANGTGYFYVLA 420
Db 573 SASGKSFEQCEAFNGYNHSTNRLTLAVAVPKYSGVSPDRCKLICRANGTGYFYVLA 632
QY 421 PKVVDGTLCSPDSTSVCGQKCIKAGCDGNGLSKKRFDKCGVCGGDNKSKCKVTGLFTRP 480
Db 633 PKVVDGTLCSPDSTSVCGQKCIKAGCDGNGLSKKRFDKCGVCGGDNKSKCKVTGLFTRP 692
QY 481 MHGYNFVAIPAGASSIDIRQRYKGLIGDNNYALKNSQKYLNGHFVYSAVERDLVV 540
Db 693 MHGYNFVAIPAGASSIDIRQRYKGLIGDNNYALKNSQKYLNGHFVYSAVERDLVV 752
QY 541 KGSLLRYSGTGTAVESLQASRPILFVEVLSVGMKTPPRVRYSFYLPKPREDKSSHP 600
Db 753 KGSLLRYSGTGTAVESLQASRPILFVEVLSVGMKTPPRVRYSFYLPKPREDKSSHP 812
QY 601 KDPGRPSVLHNSVLSNQVEQDDPPAPKRVAGSWGPCASGSGGLQKRAVDCRSAGQ 660
Db 813 KDPGRPSVLHNSVLSNQVEQDDPPAPKRVAGSWGPCASGSGGLQKRAVDCRSAGQ 872
QY 661 RTVPACDAARHPVETQACGECPTWELSAWSPCSKSGRGFORSLKCVGHGGRLLARDQ 720
Db 873 RTVPACDAARHPVETQACGECPTWELSAWSPCSKSGRGFORSLKCVGHGGRLLARDQ 932
QY 721 CNLHRKQELDFCVLRPC 738
Db 933 CNLHRKQELDFCVLRPC 950

RESULT 2
US-10-163-316-2
; Sequence 2, Application US/10163316
; Publication No. US2002019703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: MP101-025P1RNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2

Query Match 77.6%; Score 3136.5; DB 9; Length 823;
Best Local Similarity 96.2%; Pred. No. 3.7e-221;
Matches 581; Conservative 0; Mismatches 0; Indels 23; Gaps 1;
QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVKVLLL 60
Db 213 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVKVLLL 272
QY 61 RDRDGPVKTGNAALTALRNFCQAKQKLNKVSQKHEPYWDTAILFTRODLCGATTCDTLGM 120
Db 273 RDRDGPVKTGNAALTALRNFCQAKQKLNKVSQKHEPYWDTAILFTRODLCGATTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEYFGKLRANHMSP 180
Db 333 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEYFGKLRANHMSP 392

QY 181 TLIIQIDRANPWSACSAAIITDFLDGSHGDCLLDQPSKDISLPEDIPGASYTLSSQCELA 240
Db 393 TLIIQIDRANPWSACSAAIITDFLDGSHGDCLLDQPSKDISLPEDIPGASYTLSSQCELA 452
QY 241 GVGSRPCPYMQYCTKLCTGAKGQWVCOTRHFPAWADGTCGSGKCLKGACVERHNLNK 300
Db 453 GVGSRPCPYMQYCTKLCTGAKGQWVCOTRHFPAWADGTCGSGKCLKGACVERHNLNK 512
QY 301 HR-----VDGSAWAKDYPGCSRTCGGQVQLARRQCTNPTPA 337
Db 513 HRPPDIIISPKOLLRLPNGLHTTQVDGSAWAKDYPGCSRTCGGQVQLARRQCTNPTPA 572
QY 338 NGGKYCEGVVRYRSCNLEPCPSASGKSFRFEEQCEAFNGYNHSTNRLTLAVAVPKYSG 397
Db 573 NGGKYCEGVVRYRSCNLEPCPSASGKSFRFEEQCEAFNGYNHSTNRLTLAVAVPKYSG 632
QY 398 VSPDRCKLICRANGTGYFYVLA PKVVDGTLCSPDSTSVCGQKCIKAGCDGNGLSKKRF 457
Db 633 VSPDRCKLICRANGTGYFYVLA PKVVDGTLCSPDSTSVCGQKCIKAGCDGNGLSKKRF 692
QY 458 DKGCGVCGGDNKSKCKVTGLFTRKPMHGYNFVVAIPAGASSIDIRQRYKGLIGDNNYALK 517
Db 693 DKGCGVCGGDNKSKCKVTGLFTRKPMHGYNFVVAIPAGASSIDIRQRYKGLIGDNNYALK 752
QY 518 NSQKYLNGHFVYSAVERDLVVKGSLLRYSGTGTAVESLQASRPILFVEVLSVGMK 577
Db 753 NSQKYLNGHFVYSAVERDLVVKGSLLRYSGTGTAVESLQASRPILFVEVLSVGMK 812
QY 578 TTPR 581
Db 813 TTPR 816

RESULT 3
US-10-097-597-1
; Sequence 1, Application US/10097597
; Publication No. US2003002352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Ei-ji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-597-1

Query Match 55.9%; Score 2260; DB 9; Length 727;
Best Local Similarity 54.3%; Pred. No. 3.8e-157;
Matches 405; Conservative 129; Mismatches 172; Indels 40; Gaps 11;
QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVKVLLL 60
Db 13 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVKVLLL 72
QY 61 RDRDGPVKTGNAALTALRNFCQAKQKLNKVSQKHEPYWDTAILFTRODLCGATTCDTLGM 120

```

Db 73 HDQKQPEVTSNAALTLRNFNCWKQKOHNPSPDRDAEHYDTAILFTFQDLGSGQTCDTLGM 132
QY 121 ADVTGTCMPKRSQSVIEDDGLPSAFTTAHELGHVFNPHDVKVCEEVFGKLRANHMSP 180
Db 133 ADVTGTCMPKRSQSVIEDDGLQAAFTTAHELGHVFNPHDHDHAKQACASLNGVQDSHMMAS 192
QY 181 TLIOIDRANPWSACSAATITDLDGSHGDCLLDQPSKPISELPDLPFASYYTSLQOCELAF 240
Db 193 MSLNLDHSPWSPCSAYMITSLDNGHGECLMDKPNQPIQLPGDLPGTLDANRQCQFTF 252
QY 241 GVCSKPCP-YMQYCTKLWCTGAKQWVCQTRHFPWADTSGEGKCLKLGACVBERHNLN 299
Db 253 GEDSKHCPDAASTCTSLWCTGSGVLCVQTKHFPWADTSGEGKWCINGKVCNKTD-R 311
QY 300 KH---RVDGSKAWDPYPCSTCGGVQVQARRQCTNPTPANGKVCYCEGVRYKVRSCNLE 356
Db 312 KHFDTFPHGSGPWGPGWDCSRTCGGVQYVYTHRECDNPVKNGGKYCEGKRVRYSCNLE 371
QY 357 PCPSSASGKSFRQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYGF 416
Db 372 DCPDN-NGKTFREQCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLIQCAKGIFY 430
QY 417 YVLAPKVVDTGTLCPDSTSVQVQKICAGCDNGLSKRFPDKGVCGGDNKSKKVTGL 476
Db 431 FVLQPKVVDTGTLCPDSTSVQVQKICAGCDRIIDSKKFKDKGVCGGNGSTCKKISGS 490
QY 477 FTKPMGHYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNQSGKYLILNGHFVVSAYER 536
Db 491 VTSAPKGYHDVTIPTGATNIEVKQRNQRSGRNGSFLAKAADGTIILNGDYTLSTLEQ 550
QY 537 DLVVKGLLRYSGTGTAVESLQASRPILPLETVESLVSGKMTPPRVRYSFYLPKPREDK 596
Db 551 DIMYKGVWLRYSGSSAALERISFSPLEKIQTIVLGNALRPKIYTFYVKK- ---K 606
QY 597 SSHPKDPGRGPSVLHNSVLSNQVEQDDPPARVWAGWGPCSASCSGLQKRAVDCRG 656
Db 607 ES-----FNAIPTFS-----AWVIEWGECRSKSELGWORRLVECRD 643
QY 657 SAGORTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSCGRGFORRLKCVGHGG 713
Db 644 INGO---PASECAKEVKPASTRPCADHPCPQWQLGEWSSCSKTCGKYAKRSLKLSHDG 700
QY 714 RLLARQDCNLRKPKQ- LDFCVLRPC 738
Db 701 GVLSHESCDPLKPKHFIDFCTLTQC 726

```

RESULT 4

```

US-10-097-580-1
; Sequence 1, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiji
; APPLICANT: Hakozaiki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ. ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT

```

; ORGANISM: Homo sapiens
US-10-097-580-1

Query Match 55.9%; Score 2260; DB 9; Length 727;
Best Local Similarity 54.3%; Pred. No. 3.8e-157;
Matches 405; Conservative 129; Mismatches 172; Indels 40; Gaps 11;

```

QY 1 FVSPRVYETLVVADESVMVKFHGADLHYLLTLATLAARLYRHPHSILNPINIVVVKVLL 60
Db 13 FVSPRVYETMLVADQSMAEFHSGLKHYLLTLFSAARLYKHPHSIRNSVSLVVKVILVI 72
QY 61 RDRSDQPKVKTGNAALTLRNFCAWKLNKVSDEKHEPYWDYTAIFTRQDLGAGTCTLGM 120
Db 73 HDEQKGPVTSNAALTLRNFCAWKQHPSPDRDAEHYDTAILFTFQDLGSGQTCDTLGM 132
QY 121 ADVGTMCDPKRSQSVIEDDGLPSAFTTAHELGHVFNPHDVKVCEEVFGKLRANHMSP 180
Db 133 ADVTGTCMPKRSQSVIEDDGLQAAFTTAHELGHVFNPHDHDHAKQACASLNGVQDSHMMAS 192
QY 181 TLIOIDRANPWSACSAATITDLDGSHGDCLLDQPSKPISELPDLPFASYYTSLQOCELAF 240
Db 193 MSLNLDHSPWSPCSAYMITSLDNGHGECLMDKPNQPIQLPGDLPGTLDANRQCQFTF 252
QY 241 GVCSKPCP-YMQYCTKLWCTGAKQWVCQTRHFPWADTSGEGKCLKLGACVBERHNLN 299
Db 253 GEDSKHCPDAASTCTSLWCTGSGVLCVQTKHFPWADTSGEGKWCINGKVCNKTD-R 311
QY 300 KH---RVDGSKAWDPYPCSTCGGVQVQARRQCTNPTPANGKVCYCEGVRYKVRSCNLE 356
Db 312 KHFDTFPHGSGPWGPGWDCSRTCGGVQYVYTHRECDNPVKNGGKYCEGKRVRYSCNLE 371
QY 357 PCPSSASGKSFRQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYGF 416
Db 372 DCPDN-NGKTFREQCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLIQCAKGIFY 430
QY 417 YVLAPKVVDTGTLCPDSTSVQVQKICAGCDNGLSKRFPDKGVCGGDNKSKKVTGL 476
Db 431 FVLQPKVVDTGTLCPDSTSVQVQKICAGCDRIIDSKKFKDKGVCGGNGSTCKKISGS 490
QY 477 FTKPMGHYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNQSGKYLILNGHFVVSAYER 536
Db 491 VTSAPKGYHDVTIPTGATNIEVKQRNQRSGRNGSFLAKAADGTIILNGDYTLSTLEQ 550
QY 537 DLVVKGLLRYSGTGTAVESLQASRPILPLETVESLVSGKMTPPRVRYSFYLPKPREDK 596
Db 551 DIMYKGVWLRYSGSSAALERISFSPLEKIQTIVLGNALRPKIYTFYVKK- ---K 606
QY 597 SSHPKDPGRGPSVLHNSVLSNQVEQDDPPARVWAGWGPCSASCSGLQKRAVDCRG 656
Db 607 ES-----FNAIPTFS-----AWVIEWGECRSKSELGWORRLVECRD 643
QY 657 SAGORTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSCGRGFORRLKCVGHGG 713
Db 644 INGO---PASECAKEVKPASTRPCADHPCPQWQLGEWSSCSKTCGKYAKRSLKLSHDG 700
QY 714 RLLARQDCNLRKPKQ- LDFCVLRPC 738
Db 701 GVLSHESCDPLKPKHFIDFCTLTQC 726

```

RESULT 5

```

US-09-445-023A-1
; Sequence 1, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiji
; APPLICANT: Hakozaiki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical

```



```
Db 609 -----FNAIPTFS-----EWIEEWGECSTKCGSQWRVWQCRD 643
QY 657 SAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCKSCGGRGFORSLKCVGHGG 713
Db 644 INGH--PASECAKEVKPASTRPCADLPCPHWQVGDWSPCKSCGGRGFORSLKCVSHDG 700
QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
Db 701 GVLSNESCDPLKPKHYIDFCTLTQC 726

RESULT 7
US-10-097-580-12
; Sequence 12, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-097-580-12
```

```
Query Match 55.9%; Score 2259; DB 9; Length 727;
Best Local Similarity 53.6%; Pred. No. 4.5e-157;
Matches 400; Conservative 131; Mismatches 175; Indels 40; Gaps 10;

QY 1 FVSPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTILNINIVVKVLLL 60
Db 13 FVSPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTILNINIVVKVLLI 72
QY 61 RDRSGPKVTGNAALTLRNFCQAWQKLNKVSDEKHPYWDYTAILFTRODLCGATCTDGLM 120
Db 73 YEEQKGPVTSNAALTLRNFCQAWQKLNKVSDEKHPYWDYTAILFTRODLCGATCTDGLM 132
QY 121 ADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNPKVCEVFGKLRANHMSP 180
Db 133 ADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNPKVCEVFGKLRANHMSP 192
QY 181 TLIQIDRANPWSACSAIITDLSGHDCLLDQPSKPISLPEDLPASGATLSQOCELA 240
Db 193 MLSSLDHSQWSPSCSAIITDLSGHDCLLDQPSKPISLPEDLPASGATLSQOCELA 252
QY 241 GVSKPCP-YMQYCTKLWCTGKAGQWCVQTRHPFWADGTSCGEGKLCACVERHNLN 299
Db 253 GEESKPCP-YMQYCTKLWCTGKAGQWCVQTRHPFWADGTSCGEGKLCACVERHNLN 311
QY 300 KH---RVDGSAKWDYGPSCSRTCGGQVQLARROCTNPTPANGSKYCEGVRYKRNLE 356
Db 312 KHFPATPVHSGWGPWGPWGDSCRTCGGQVQYTWRECDNPVPNGKYGKCEGARVIRSCNIE 371
QY 357 PCPSASGKSPREQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKKCLICRANGTYF 416
Db 372 DCPDN-NGKTFREBQCEAHNEFSKASFGNEPTWEVTPKYAGVSPKDKCLTCEAKGIGYF 430
```

```
QY 417 YVLAPKVVDTGLTSPDSTSVCGQKCIKAGCDGNLGSKKRDKCGVCGGDNKSKKVTGL 476
Db 431 FVLQPKVVDTGLTSPDSTSVCGQKCIKAGCDRIIDSKKFKDCGCGGNGSTCKKMSGI 490
QY 477 FTKPMHGYNVFAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNGHFEVWSAVER 536
Db 491 VTSTRPGYHDIVTPAGATNIEVKHRNQRNGSRNNGSFLAIRAADGTIYILNGFTLTLEQ 550
QY 537 DLVVGSLRLRYSGTGTAVESIOASRPILTEPLTVELSVSKMTPTPRVYSFYLPKEPREDK 596
Db 551 DLYTKGTVLRYSGSSAALERIRSFPLKEPLTIQVLMVGHALRPKIKFTFYFMKKKTES-- 608
QY 597 SSHPKDPRGSPVLHNSVLSLSNQVEQPDPRPPARWAGVSGWGPCSASCSGSLQKRAYDCRG 656
Db 609 -----FNAIPTFS-----EWIEEWGECSTKCGSQWRVWQCRD 643
QY 657 SAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCKSCGGRGFORSLKCVGHGG 713
Db 644 INGH--PASECAKEVKPASTRPCADLPCPHWQVGDWSPCKSCGGRGFORSLKCVSHDG 700
QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
Db 701 GVLSNESCDPLKPKHYIDFCTLTQC 726

RESULT 8
US-09-445-023A-12
; Sequence 12, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/09/445,023A
; CURRENT FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-445-023A-12

Query Match 55.9%; Score 2259; DB 10; Length 727;
Best Local Similarity 53.6%; Pred. No. 4.5e-157;
Matches 400; Conservative 131; Mismatches 175; Indels 40; Gaps 10;

QY 1 FVSPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTILNINIVVKVLLL 60
Db 13 FVSPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTILNINIVVKVLLI 72
QY 61 RDRSGPKVTGNAALTLRNFCQAWQKLNKVSDEKHPYWDYTAILFTRODLCGATCTDGLM 120
Db 73 YEEQKGPVTSNAALTLRNFCQAWQKLNKVSDEKHPYWDYTAILFTRODLCGATCTDGLM 132
QY 121 ADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNPKVCEVFGKLRANHMSP 180
Db 133 ADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNPKVCEVFGKLRANHMSP 192
QY 181 TLIQIDRANPWSACSAIITDLSGHDCLLDQPSKPISLPEDLPASGATLSQOCELA 240
Db 193 MLSSLDHSQWSPSCSAIITDLSGHDCLLDQPSKPISLPEDLPASGATLSQOCELA 252
QY 241 GVSKPCP-YMQYCTKLWCTGKAGQWCVQTRHPFWADGTSCGEGKLCACVERHNLN 299
```

Db 357 ADVGTVCDPSCSVIEDDGLQAAFTTAHELGHVENPHDDAKKASLNGVTDGSHLMAS 416
 QY 181 TLTOIDRANPWSACSAIITDFDSDGHDCLLOPQSPKIPSLPEDLPASVTLTSCQCELAF 240
 Db 417 MLSSLDHSPWPCSAAYMTSFLDNGHGECLMDKPNKIPKLSPLDPTLYDANRQCOFTF 476
 QY 241 GVGSKPCP-YMQYCTKLWCTGKAGOMVCTRHFPWADGTSCGEGKCLKAGACVERHNLN 299
 Db 477 GEESKHCFDAASTCTTLMWCTGTSGLLVCOTKHFPWADGTSCGEGKWCVCVKNTDM- 535
 QY 300 KH---RYDGSWAKWDYPGCSRTCGGVQLARRQCTNPTPANGKCYCEGVRYVYRSCNLE 356
 Db 536 KHFAIPVHSGWPGMGDCSRTCGGVQYTMRECDNPVPKNGKCYCEGVRYVYRSCNIE 595
 QY 357 PCPSASGKSPREOCBAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYGF 416
 Db 596 DCPDN-NGKTPREOCBAHNEFSKAFNEPTVEWTKYAGVSPKDRCKLTCEAKGIGYF 654
 QY 417 YVLAPKVVDGTLCPDSTSVCOGKCIKAGDGNLGSKRFDKCGVCGGDKCKKVTGL 476
 Db 655 FVLQPKVVDGTPCSPDSTSVCOGCVKAGCDRIIDSKKFKDCCGCGGNGSTCKKMSGI 714
 QY 477 FTKPMHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNQSGKYLLNGHVFVSAVER 536
 Db 715 VTSTRPGYHDIVTIPAGATNIEVHRNQRGSRNNGSFLAIRADGTYILNGFTLSTLEQ 774
 QY 537 DLVWKSLLRYSGTGTAVESLOASRPILPTVEVLSVGKMTPPRVRYSFYLPKPREDK 596
 Db 775 DLTKYKTVLRYSGSAALEIRISFSLKEPLTIQVLMVGHALRPKIKFTYFMKKKTES-- 832
 QY 597 SSHFKDPRGSPVJHNSVLSLNOVEQDDPRPARWAGSMGSPCSASGSGSLQKRAVDCRG 656
 Db 833 -----FNAIPTFS-----EWVIEEWGECSTCGSGWQRRVQCRD 867
 QY 657 SAGQRTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSCGRGFORSLKCVGHGG 713
 Db 868 INGH---PASECAKEVKPASTRPCADLPCPHWQVGDWSPCSKTCGKGYKRTLCVSHDG 924
 QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
 Db 925 GVLNSCEDPLKPKHYIDFCTLTQC 950
 RESULT 10
 US-10-163-316-7
 ; Sequence 7, Application US/10163316
 ; Publication No. US20020197703A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kapeller-Libermann, Rosana
 ; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
 ; FILE REFERENCE: MP101-025PFRNN
 ; CURRENT APPLICATION NUMBER: US/10/163,316
 ; PRIOR FILING DATE: 2002-06-05
 ; PRIOR APPLICATION NUMBER: 60/297,863
 ; PRIOR FILING DATE: 2001-06-13
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 968
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-163-316-7
 Query Match 55.8%; Score 2257; DB 9; Length 968;
 Best Local Similarity 53.6%; Pred. No. 9e-157;
 Matches 400; Conservative 130; Mismatches 176; Indels 40; Gaps 10;
 QY 1 FVSIPRVYETLVVADESVMVKFGADLEHYLLTLATARLYRHPISLNPINIVVVKVLL 60
 Db 254 FVSPRVYETLVVADESVMVKFGADLEHYLLTLATARLYRHPISLNPINIVVVKVLL 313
 QY 61 RDRSGPKVTGNAALTLNFCNAWQKLNKVSDEKHPYWDYLTFTQDLGGATTCDTLGM 120

Db 253 GEESKHCPDAASTCTTLMWCTGTSGLLVCOTKHFPWADGTSCGEGKWCVCVKNTDM- 311
 QY 300 KH---RYDGSWAKWDYPGCSRTCGGVQLARRQCTNPTPANGKCYCEGVRYVYRSCNLE 356
 Db 312 KHFAIPVHSGWPGMGDCSRTCGGVQYTMRECDNPVPKNGKCYCEGVRYVYRSCNIE 371
 QY 357 PCPSASGKSPREOCBAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYGF 416
 Db 430 DCPDN-NGKTPREOCBAHNEFSKAFNEPTVEWTKYAGVSPKDRCKLTCEAKGIGYF 430
 QY 417 YVLAPKVVDGTLCPDSTSVCOGKCIKAGDGNLGSKRFDKCGVCGGDKCKKVTGL 476
 Db 490 FVLQPKVVDGTPCSPDSTSVCOGCVKAGCDRIIDSKKFKDCCGCGGNGSTCKKMSGI 490
 QY 477 FTKPMHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNQSGKYLLNGHVFVSAVER 536
 Db 550 VTSTRPGYHDIVTIPAGATNIEVHRNQRGSRNNGSFLAIRADGTYILNGFTLSTLEQ 550
 QY 537 DLVWKSLLRYSGTGTAVESLOASRPILPTVEVLSVGKMTPPRVRYSFYLPKPREDK 596
 Db 608 DLTKYKTVLRYSGSAALEIRISFSLKEPLTIQVLMVGHALRPKIKFTYFMKKKTES-- 608
 QY 597 SSHFKDPRGSPVJHNSVLSLNOVEQDDPRPARWAGSMGSPCSASGSGSLQKRAVDCRG 656
 Db 643 -----FNAIPTFS-----EWVIEEWGECSTCGSGWQRRVQCRD 643
 QY 657 SAGQRTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSCGRGFORSLKCVGHGG 713
 Db 701 GVLNSCEDPLKPKHYIDFCTLTQC 726
 QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
 Db 726 GVLNSCEDPLKPKHYIDFCTLTQC 726
 RESULT 9
 US-09-321-987B-4
 ; Sequence 4, Application US/09321987B
 ; Patent No. US20020102210A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kimble, Judith E
 ; APPLICANT: Belloch, Robert H
 ; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
 ; FILE REFERENCE: 960296.95386
 ; CURRENT APPLICATION NUMBER: US/09/321,987B
 ; CURRENT FILING DATE: 1999-05-28
 ; PRIOR APPLICATION NUMBER: 60/087,170
 ; PRIOR FILING DATE: 1998-05-29
 ; PRIOR APPLICATION NUMBER: 60/129,023
 ; PRIOR FILING DATE: 1999-04-13
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 950
 ; TYPE: PRT
 ; ORGANISM: Murine
 US-09-321-987B-4
 Query Match 55.8%; Score 2257; DB 10; Length 950;
 Best Local Similarity 53.6%; Pred. No. 8.8e-157;
 Matches 400; Conservative 130; Mismatches 176; Indels 40; Gaps 10;
 QY 1 FVSIPRVYETLVVADESVMVKFGADLEHYLLTLATARLYRHPISLNPINIVVVKVLL 60
 Db 237 FVSPRVYETLVVADESVMVKFGADLEHYLLTLATARLYRHPISLNPINIVVVKVLL 296
 QY 61 RDRSGPKVTGNAALTLNFCNAWQKLNKVSDEKHPYWDYLTFTQDLGGATTCDTLGM 120
 Db 257 YEEQKPEVTGNAALTLNFCNAWQKLNKVSDEKHPYWDYLTFTQDLGGATTCDTLGM 356
 QY 121 ADVGTMCDPKSCSVIEDDGLPAAFTTAHELGHVENPHDDAKKASLNGVTDGSHLMAS 180

```
Db 314 YEQKGPVTSNAALTRNFCWQKNSPSDRDPEHYDTAILTRDGLCSHFTCDLGM 373
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNDVNVKCEEVFGKLRANHMSP 180
Db 374 ADVGTVCDPKRSRSCSVIEDDGLQAAFTTAHELGHVFNPHDNDVNVKCEEVFGKLRANHMSP 433
QY 181 TLQIDRANPWSACSAIIITDFDSGHGCDLLDQSPKIPSLDEPLPGASYTLSQOCELAF 240
Db 434 MLSSLDHSQPSWPCSAIYMTSFLDNGHGBCLMDKPNQIKLPDLPGTLYDANRQCQFTF 493
QY 241 GVGSKPCP-YMOYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKCLKACVERHNIN 299
Db 494 GEESKPCDAASTCTLWCTGSGLLVCQTKHFPWADGTSCGEGKCLKACVERHNIN 552
QY 300 KH---RVDSWAKWDYPGCSRTCGGVQVQLARRQCTNPTPANGKYCEGVRYKRSCLNLE 356
Db 553 KHFATPVHSGWPGWGPCSRTCGGVQVYTMRECDNPNVKNKGKCEGVRYKRSCLNLE 612
QY 357 PCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSGVSPRDKCKLICRANGTYGF 416
Db 613 DCPDN-NGKTFREOCEAHNEFSKASFCNEPTVEWTPKYAGVSPKDRCKLTCEAKGIGYF 671
QY 417 YVLAPKVVVDGTLCSPTDSTSVCGQKICAGQDGNLGSKKRFDKCGVCGDNKCKKVTGL 476
Db 672 FVLQPKVVVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCKKMSGI 731
QY 477 FTKPMHGVNFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNGHFVVSAYER 536
Db 732 VTSTRPGVHDIVTIPAGATNIEVKQRNQRSGNSFLAIRAADGTYLNGNTFTLSLEQ 791
QY 537 DLVVKGLLRYSYGTGTAVERLOASRPILPLTEVLVSQKMTPPRVKYSYLPKPEPREDK 596
Db 792 DLTKGVILRVYSGSSAALERISFSPKLEPLTQVLTVGNALRPKIKITYFYVKKK---K 846
QY 597 SSHPKDPRGPSVNLHNSVLNSQVEQDPPRPPARWAGSWGSPCSASGSLQKRAVDCRG 656
Db 850 -----FNAIPTFS-----AWVIEEWGECSSKCELGWQRRLVECDR 883
QY 657 SAGQRTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSGRGFORRLKCVGHGG 713
Db 885 INGH---PASECAKEVPKASTPCADLPCPHWQVGDWSPCSKTCGKGYKRLKCLSHDG 941
QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
Db 942 GVLNESCDELKPKKHIDFCTLTQC 967
```

RESULT 11

```
US-10-105-929-2
; Sequence 2, Application US/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 0940/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-105-929-2
```

```
Query Match 55.8%; Score 2254; DB 12; Length 967;
Best Local Similarity 54.0%; Pred. No. 1.5e-156;
Matches 403; Conservative 131; Mismatches 172; Indels 40; Gaps 11;

QY 1 FVSIPIRYVETLVVADSWKFKGADLEHYLLTLLATAARLYRHPSTILNINIVVKKVLL 60
Db 253 FVSHRYVETLVVADSWKFKGADLEHYLLTLLATAARLYRHPSTILNINIVVKKVLL 312
QY 61 RDRSDGPKVTGNAALTRNFCWQKNSPSDRDPEHYDTAILTRDGLCSHFTCDLGM 120
Db 313 HDQKQGEVTSNAALTRNFCWQKNSPSDRDPEHYDTAILTRDGLCSHFTCDLGM 372
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNDVNVKCEEVFGKLRANHMSP 180
Db 373 ADVGTVCDPKRSRSCSVIEDDGLQAAFTTAHELGHVFNPHDNDVNVKCEEVFGKLRANHMSP 432
QY 181 TLQIDRANPWSACSAIIITDFDSGHGCDLLDQSPKIPSLDEPLPGASYTLSQOCELAF 240
Db 433 MLSSLDHSQPSWPCSAIYMTSFLDNGHGBCLMDKPNQIKLPDLPGTLYDANRQCQFTF 492
QY 241 GVGSKPCP-YMOYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKCLKACVERHNIN 299
Db 493 GEESKPCDAASTCTLWCTGSGLLVCQTKHFPWADGTSCGEGKCLKACVERHNIN 551
QY 300 KH---RVDSWAKWDYPGCSRTCGGVQVQLARRQCTNPTPANGKYCEGVRYKRSCLNLE 356
Db 552 KHFATPVHSGWPGWGPCSRTCGGVQVYTMRECDNPNVKNKGKCEGVRYKRSCLNLE 611
QY 357 PCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSGVSPRDKCKLICRANGTYGF 416
Db 612 DCPDN-NGKTFREOCEAHNEFSKASFCNEPTVEWTPKYAGVSPKDRCKLTCEAKGIGYF 670
QY 417 YVLAPKVVVDGTLCSPTDSTSVCGQKICAGQDGNLGSKKRFDKCGVCGDNKCKKVTGL 476
Db 671 FVLQPKVVVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCKKISG 730
QY 477 FTKPMHGVNFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNGHFVVSAYER 536
Db 731 VTSAPKGVHDIVTIPAGATNIEVKQRNQRSGNSFLAIRAADGTYLNGNTFTLSLEQ 790
QY 537 DLVVKGLLRYSYGTGTAVERLOASRPILPLTEVLVSQKMTPPRVKYSYLPKPEPREDK 596
Db 791 DLTKGVILRVYSGSSAALERISFSPKLEPLTQVLTVGNALRPKIKITYFYVKKK---K 846
QY 597 SSHPKDPRGPSVNLHNSVLNSQVEQDPPRPPARWAGSWGSPCSASGSLQKRAVDCRG 656
Db 847 ES-----FNAIPTFS-----AWVIEEWGECSSKCELGWQRRLVECDR 883
QY 657 SAGQRTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSGRGFORRLKCVGHGG 713
Db 884 INGH---PASECAKEVPKASTPCADLPCPHWQVGDWSPCSKTCGKGYKRLKCLSHDG 940
QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
Db 941 GVLNESCDELKPKKHIDFCTMAEC 966
```

RESULT 12

```
US-09-918-171A-9
; Sequence 9, Application US/09918171A
; Patent No. US20020110894A1
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/04193
; CURRENT APPLICATION NUMBER: US/09/918,171A
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 09/369,364
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 9
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-8
US-09-918-171A-9

Query Match          47.3%; Score 1911.5; DB 10; Length 905;
Best Local Similarity 50.6%; Pred. No. 1.5e-131;
Matches 352; Conservative 127; Mismatches 185; Indels 31; Gaps 13;

QY 1 FVSIPRYETLVVADSMVKFHGADLEHYLLTLLATAARLYRHPSTLNPINIVVVKVLL 60
DB 229 FVSEARFETLVADASMAAFYGTDLQNHILTVMSMAARIYKHPSTRNSVNLVVKVLI 288
QY 61 RDRDSGPKVTGNALTLNFCNAWKLNKVDKHPYWDTAILFTRODLG-ATTCDTLG 119
DB 289 EKERWGEVSDNGGLTLNFCNSWRFRNPKPSDRHPHYDTALFTKQFCGGEODTLG 348
QY 120 MADVGTMCOPKRCSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMS 179
DB 349 MADVGTICDPKSCSVIKDEGLQAAYTLAHELGHVLSMPHDDSKPCVRLFGPMGVHMA 408
QY 180 PTLIQIDRANPWSACSAIITDLDGSHGDCILLDQPSKIPISLPEDLPAS--YTLSSQCE 237
DB 409 PFTIHNKTLTPWSPCSAVYLTILLDGHGDCILLDAPTSVLPFTGLPGHSTLYELDQQCK 468
QY 238 LAFVGSKPCP---YMQYCTKLWCTKAKGQWVCOTRH--FWADGTSCGEGKCLKLGAC 292
DB 469 QIFGPDPRHCNTSVEDICVQLCARHDSDEPICHTKNGLSLWADGTCGPGHCLDGS 528
QY 293 VERHNLN--KHRYDGSNAKWDYPCSRCTCGGVQLARQCTNPTPANGKGYCEGVRYKY 350
DB 529 VLKEDVENKAVVDGWPMPWPGQCSRTCGGIGIQPSNRECDNPMQNGRFGCLGERVY 588
QY 351 RSNLEPCPSASGKSFREOCEAFNGYHNSTNRLTLAVWPVKYSGVSPROCKLICRA 410
DB 589 QSCNTEEC--PNKGKSFREQCEKYNAYNH--TDLGNEFLQWVPKYSGVSPROCKLFCRA 645
QY 411 NGTYFVVLAPKVVVDGLTCLSPDSTSVCGVKCIKAGCDNLGSKKRFDCGCGGDNKSC 470
DB 646 RGRSEKVFPEAKVIDGTLCGPDFTLSICVRQCVKAGCDHVNPSPKLDRKCGVGGGTAC 705
QY 471 KKVYGLTKPMHGYNFVVAIPAGASSIDIRQYKGLIGDNDYLLAKNSGKTYLLNGHFV 530
DB 706 RKISGFTPEYSGYNDIVTIPAGATNIDVKQRSHPGVYRNDGSYLALKTANGQYLLNGNLA 765
QY 531 VSAVERDLVKGSLRYSGTGTAVESIQASRPILPTVEVLSV-GKMTPPRVRYSFYLP 589
DB 766 ISATEQDILVKGTLTKYSGSMATLERIQSFQALPELTVQLLTVSGEVFPKRYTFYFP 825
QY 590 KEPREDKSHHPKDPKRGPSVLHNSVLSLNQVEQDPRPPARWVAGSWGPCASCSGLQK 649
DB 826 ND--MDFSVQNSKERAT--NIIQSL-----PSAEWVLGDWSECPSTCRGSMQR 870
QY 650 RAVDCRGSAGQRTVPACDAARHPVETOACG-EPCC 683
DB 871 RTVECDRPSGQAS-DTCDEALKPEDAKPCGSPCP 904

RESULT 13
US-09-803-589-2
; Sequence 2, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803, 589
; CURRENT FILING DATE: 2001-03-09
```

```
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-803-589-2

Query Match          47.0%; Score 1901; DB 10; Length 608;
Best Local Similarity 49.3%; Pred. No. 5.3e-131;
Matches 339; Conservative 114; Mismatches 146; Indels 88; Gaps 8;

QY 57 VLLLRDRSGPKVTGNALTLNFCNAWKLNKVDKHPYWDTAILFTRODLGATTC 116
DB 4 ILVIHDEQKPEVTSNAALTNRNFCNQKQHNPPSDRDAEHYDTAILFTRODLGSGTCD 63
QY 117 TLGMADVTCMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANH 176
DB 64 TLGMADVTCMDPKRSCSVIEDDGLQAFAFTTAHELGHVFNPHDDAKOCASLNGVNDSH 123
QY 177 MMSPTLIQIDRANPWSACSAIITDLDGSHGDCILLDQPSKIPISLPEDLPASGTYLSSQ 236
DB 124 MMSMLNLDHISQWSPCSAYNITSFLDNGHECLMDKDPONPIQLPGDLPTSYDANRQC 183
QY 237 ELAFVGSKPCP-YMQYCTKLWCTKAKGQWVCOTRHFPWADGTSCGEGKCLKLGACVER 295
DB 184 QFTFEGSKHCPDAASTCTSLMCTGTSGVLVCQTQKHPWADGTSCGEGKWCINGKCVNK 243
QY 296 HNLNKH---RVDGSNAKWDYPCSRCTCGGVQLARQCTNPTPANGKGYCEGVRYKYRS 352
DB 244 TD-RKHEFTPPHSGWGMGMPGDCSRTCGGVQYTMRECDNPPVKNKGKCYCEKRVYRS 302
QY 353 CNLEPCPSASGKSFREOCEAFNGYHNSTNRLTLAVWPVKYSGVSPRDKCLICRANG 412
DB 303 CNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLICQAKG 361
QY 413 TGTYFVVLAPKVVVDGLTCLSPDSTSVCGVKCIKAGCDNLGSKKRFDCGCGGDNKSCCK 472
DB 362 IGYFFVLOPKVVVDGTPCSPDSTSVCGVQCQKAGCDRIIDSKKFKDKCGVCGGNGSTCKK 421
QY 473 VTGLFTKPMHGYNFVVAIPAGASSIDIRQYKGLIGDNDYLLAKNSGKTYLLNGHFVYS 532
DB 422 ISGVSISAQPGYHDIITPTGATNIEVKQRNORSGRNSGFLAIKAADGYTLNGDYTLS 481
QY 533 AVERDLVKGSLRYSGTGTAVESIQASRPILPTVEVLSVGVKMTPPRVRYSFYLPKEP 592
DB 482 TLEQDIMYKGVVLYSGSSAALIRSFSPLEPTLIQVLTVGNALRPKIRKYTFYEVKKK- 540
QY 593 REDKSSHHPKDPKRGPSVLHNSVLSLNQVEQDPRPPARWVAGSWGPCASCSGGLQKRAV 652
DB 541 ---KES-----FNAITFS----- 551
QY 653 DCRGSAGQRTVPACDAARHPVETOACGEPCTWELSAWSPCSKSGRGFRRLSKCVGHG 712
DB 552 -----AWIIEGECCKTCTGKGYKKRSLKCLSHD 580
QY 713 GRLLARDQCNLHRKPOE-LDFCVLRPC 738
DB 581 GGVLSHESCDPLKPKHFDICTMAEC 607
```

RESULT 14
US-09-803-589-8
; Sequence 8, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-803-589-8

Query Match 47.0%; Score 1900; DB 10; Length 608;
Best Local Similarity 49.3%; Pred. No. 6.3e-131;
Matches 339; Conservative 114; Mismatches 146; Indels 88; Gaps 8;
QY 57 VLLLRDRSGPKVTGNAALTNRNFCWQKLNKVSXKHPYWDYTAIFTRQDLGATTC 116
DB 4 ILVTHDEQKGPVTSNAALTNRNFCWQKHNPPSDRDAEHYDTAIFTRQDLGATTC 63
QY 117 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEVFGKLRANH 176
DB 64 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEVFGKLRANH 123
QY 177 MMSPTLIQIDRANPWSACSAAITDFLDGSHGDCILDDQSPKIPSLPDLPGASYTLGSOQC 236
DB 124 MMSPTLIQIDRANPWSACSAAITDFLDGSHGDCILDDQSPKIPSLPDLPGASYTLGSOQC 183
QY 237 ELAFGVGSKPCP-YMYCTKLMCTGKAGQWVCQTRHFPWADGTSCEGKCLKLGACVYR 295
DB 184 QFTFGEDSKHCPDAASTCTLWCTGTSGVLVCQTKHFPWADGTSCEGKCLKLGACVYR 243
QY 296 HNLNKH---RVDSNAKWDPYGCSTRTCGGVLQARRQCTNPTPANGKYCEGVYRVYRS 352
DB 244 TD-RKHFDTPFHGSMGMWPGDCSRTCGGVQYTMRECDNPVKNKKYCEGVYRVYRS 302
QY 353 CNLEPCSSASGKSFREEQCEAFNGYNHSTNRLTLAVAVPKYSGVSPDKCLICRANG 412
DB 303 CNLEPCSSASGKSFREEQCEAFNGYNHSTNRLTLAVAVPKYSGVSPDKCLICRANG 361
QY 413 TGYFVVLAPKVVDTLCSPDSTSVCVQKCIKAGDGNLGSKRFDCGVCNGDNKSKCK 472
DB 362 IGYFFVLQPKVVDGTPCSPDSTSVCVQKCIKAGDGNLGSKRFDCGVCNGDNKSKCK 421
QY 473 VTGLFTKPMHGVNFVAIPAGASSIDIRQGYKGLIGDNNYLAALKNSQKYLINGHFVYS 532
DB 422 IGSVTSAPKPGYHDIITIPIGATNIEVKQRNQRNNGSFLAKAADDYTLINGDYTLIS 481

QY 533 AVERDLVVGSLRLRYSGTGTAVERSLQASRPILPLTVEVLSVGKMTPPRVRYSYFLPKPEP 592
DB 482 TLEODIMYKGVVLRYSAGSAAERIRSFPEKEPLTIQVLTGNALRPLKITYFYFKKK- 540
QY 593 REDKSSHPKDPGRGVSFLHNSVLSLSNQVEQDDPPARWVAGSWGPCASCSGLQKRAV 652
DB 541 ---KES-----FNAIPTFS----- 551
QY 653 DCRGSAGQRTVPACDAAHRPVETQACGEPCTWELSAWSPCSKSGRGFORRLKCVGHG 712
DB 552 -----AWVIEEWGECSTCKGKGYKRSKLCLSHD 580
QY 713 GRLIARDQCNLHRKPOE-LDFCVLRPC 738
DB 581 GGVLSHESCDPLKPKHFIDFCTMAEC 607
RESULT 15
US-10-105-929-13
; Sequence 13, Application US/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-105-929-13

Query Match 47.0%; Score 1900; DB 12; Length 608;
Best Local Similarity 49.3%; Pred. No. 6.3e-131;
Matches 339; Conservative 114; Mismatches 146; Indels 88; Gaps 8;
QY 57 VLLLRDRSGPKVTGNAALTNRNFCWQKLNKVSXKHPYWDYTAIFTRQDLGATTC 116
DB 4 ILVTHDEQKGPVTSNAALTNRNFCWQKHNPPSDRDAEHYDTAIFTRQDLGATTC 63
QY 117 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEVFGKLRANH 176
DB 64 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEVFGKLRANH 123
QY 177 MMSPTLIQIDRANPWSACSAAITDFLDGSHGDCILDDQSPKIPSLPDLPGASYTLGSOQC 236
DB 124 MMSPTLIQIDRANPWSACSAAITDFLDGSHGDCILDDQSPKIPSLPDLPGASYTLGSOQC 183
QY 237 ELAFGVGSKPCP-YMYCTKLMCTGKAGQWVCQTRHFPWADGTSCEGKCLKLGACVYR 295
DB 184 QFTFGEDSKHCPDAASTCTLWCTGTSGVLVCQTKHFPWADGTSCEGKCLKLGACVYR 243
QY 296 HNLNKH---RVDSNAKWDPYGCSTRTCGGVLQARRQCTNPTPANGKYCEGVYRVYRS 352
DB 244 TD-RKHFDTPFHGSMGMWPGDCSRTCGGVQYTMRECDNPVKNKKYCEGVYRVYRS 302
QY 353 CNLEPCSSASGKSFREEQCEAFNGYNHSTNRLTLAVAVPKYSGVSPDKCLICRANG 412
DB 303 CNLEPCSSASGKSFREEQCEAFNGYNHSTNRLTLAVAVPKYSGVSPDKCLICRANG 361
QY 413 TGYFVVLAPKVVDTLCSPDSTSVCVQKCIKAGDGNLGSKRFDCGVCNGDNKSKCK 472

QY 591 EPREDKSSHPKDRGPSVLHNSVLSNQVEQDDPRPARWVAGSWGSPCSASCGLQKR 650
Db 428 K-----KES-----FNAIPFTS-----AWVIEWGECSKSCGLGWQR 460
QY 651 AYDCRGSAGQRTVPACAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGFORRLK 707
Db 461 LVECRDINGQ---PASECAKEVKPASTRPCADHPCPQWLGWSESCSKYCKGKYKRLK 517
QY 708 CVGHGRLLARDQCNLHRKPQE-LDFCVLRPC 738
Db 518 CLSHDGGVLSHESCDPLKKPKHFIDFCTMAEC 549
RESULT 4
T13171
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21371; T24896
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <W1>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone F25H8
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <W2>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
C:Genetics:
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 817/1
Query Match 38.8%; Score 1567; DB 2; Length 2165;
Best Local Similarity 39.6%; Pred. No. 2.9e-99;
Matches 321; Conservative 113; Mismatches 283; Indels 94; Gaps 17;
QY 7 YETLVVADESVMKPHGADLEHYLLTLTATAARLYRHPHSILNPINIVVVKVLLLRDRDSG 66
Db 281 YVEVLVADTKMYEYHGRSLEDYVLTFTSVASIRHQSLRASINVVVVKVLLVLTENAG 340
QY 67 PKVTGNAALTNRNFCAMOKKLNKVSXDKHPEYWDTAILTFRDLCGAT-TCDTLGMADVGT 125
Db 341 PRITQNAQTLDQFCRWQOYNDPDDSSVQHDVAILTRKIDCRSQGKCDTLGLAELGT 400
QY 126 MCDPRKSCSVIBDDGLPSAFTTAHELGHVFNPHDNVVCVEV-----FGKL 172
Db 401 MCDMOKSCAILEDNGLSAFTTAHELGHVFSIPHDDEKCSYTPYVNVKVFQSKFDKT 460
QY 173 RAN----HMSPTLIIDRANPWSACSAIITFDLDSGGH--DCLLDQSKPSLPE---- 223
Db 461 QQNRFHIMAPTLEYNTHPWSNPSGAGMLERFLENNRGOTQCLFDQPVERRYEDFVR 520
QY 224 DLPGASYTLSSQCELAFGVSKPCPYMOYCTKLNCTGKAKGMVCOITHFHWADGTCGE 283
Db 521 DEFGKKYDAHQCKEYFGPASELCYPTMTCRLMWCATYFGSOMGCRTOHMPWADGTPCDE 580
QY 284 GK--LCLKGCACVERHNLRKRVDSWAKWDYPGCSRTCCGGVQLARRQCTNPTPANGGK 341
Db 581 SRSMFCHGACVRLAPESITKIDGQWDRSNWGECSRTCCGGVQKGLDQCDSPRNGGK 640
QY 342 YCGVYVYRSCNLEBPCPSSAGSKSFRFEEQCEAFNGYN-----HSTNRLTLAVAWVPKY 395
Db 342 YCGVYVYRSCNLEBPCPSSAGSKSFRFEEQCEAFNGYN-----HSTNRLTLAVAWVPKY 395

Db 641 YCVQQRERYRSCNTQECFWDT--QPYREVOCSENNKNDIGIQVASTN-----THWVPKY 693
QY 396 SGVSPRDKCLICRANGTYGYVVLAPKVDGTLCSPDSTSVYQVQKCIKAGCDNGLSKK 455
Db 694 ANVAPNERCKLYCRLSGSAAFYLLRDKVVDGTPCDNRNGDDICVAGACMPAGCDQLHSTL 753
QY 456 RFDKCGVCGGDNCKSKVYGLFKP-MHGFNFVVAIPAGASSIDIRQGYKGLIGDDNYL 514
Db 754 RRDCKGVCGGDSSCKVVKGTNEQGTGFNEYVKNKIPAGSANIDIRQGYNNKMKEDNYL 813
QY 515 ALKNSQKYLINGHFVVSAYERDLVVKGSLLRYSGTAVESLQASRPIELPTVFLSV 574
Db 814 SLRAANGEFLNGHFQVSLARQIAFDYVLEYSGSDAILIERINGTPIRSDIYVHLSV 873
QY 575 GKMTPPRVRYSFYLPKEPRE-----DKSS 598
Db 874 GS-HPPDISYEYMTAAVNAVIRPISALYLWRVTDTWTECDRACRQSQSKLCLMDMST 932
QY 599 HPKDPGPSVLHN----SVLSLSNQVEQDDPRPARWVAGSWGSPCSASCGLQKRAYDC 654
Db 933 HROS-----HNRNCONVLKPKOATRMCNIDCSTRITEDVSSCAKCGSKQRQVSC 985
QY 655 RGSAGQRTVPA---CDAARHPVETQACGPPCP--TWELSAWSPCSKSCG-RGFQRRSLK 707
Db 986 VKMGGDRQTPASEHLCDRNSKPSDIASCIYDCSGRKNWYGEWTSCTCGSNGKMKRKS 1045
QY 708 CVGHGRLLARDQCNLHRKPQELDFCVLRPC 738
Db 1046 CVDDSNRRVDESCLGRQKEATERECNRIPC 1076
RESULT 5
T18517
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine
N:Alternate names: procollagen N-proteinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18517
R:Colige, A.; Nussgens, B.V.; Lapiere, C.M.
submitted to the EMBL Data Library, February 1996
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A:Reference number: Z18941
A:Accession: T18517
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1205 <COL>
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1
A:Experimental source: skin
C:Genetics:
A:Gene: PC I-NP
C:Function:
A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior
C:Keywords: hydrolase; metalloproteinase
Query Match 25.9%; Score 1045.5; DB 2; Length 1205;
Best Local Similarity 33.7%; Pred. No. 9.6e-64;
Matches 260; Conservative 108; Mismatches 300; Indels 103; Gaps 31;
QY 8 VETLVVADESVMKPHGAD-LEHYLLTLTATAARLYRHPHSILNPINIVVVKVLLLRDRDSG 66
Db 262 IEVLGVDSDSVQFHGTEHVQKYLTLMTNVTNEIYHDES LGAHINVLVRIILSYGKSM 321
QY 67 PKV-TGNAALTNRNFCAMOKKLNKVSXDKHPEYWDTAILTFRDLCGATTCDTLGMADVGT 125
Db 322 SLIEIGNPSQSLNVCNWAYLOQRPDTHDEYHDHAIFLTRODF-GPSGMQ--GYAPVTG 378
QY 126 MCDPRKSCSVIBDDGLPSAFTTAHELGHVFNPHDNV-KVCEVFGKLNHMSPTLIQ 184
Db 379 MCHPVRCTLNHEDGSSAFVVAHETGHVLGMEDHGGNRCGD---EVLRLGSIAPLVA 435
QY 185 IDRANPWSACSAIITFDLDSGGHCDLLDQ---SKPISLPEDLPAGSYTLSSQCEAFNG 241
Db 436 AFHRFHSRCSQELSRYSLS--YDCLRDDPFTHDWP-ALPQ-LPGLHYSMNEQCRDFG 491

Db	173	ADKVVDTGKCDKSNDDICVDGECIPVCGDKGLGSSLKFDKCGDKDGGSTCKTIEGRFDE	232
Qy	480	P--MHGVNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNSOGKYLLNGHVVSAYERD	537
Db	233	RNLSPGYEDIIKLPEGATNIQEAR-----KSTNNLALKNGSDHFLYVINGLIQ-VEKE	286
Qy	538	LWVGKSLRLRYSGTAVESLOASRPILLEPTLVEVLSVGKMTPPRRVRYSFYLPKPREPRED--	595
Db	287	VEVGGTIFVYD--DAEPETLSAOGPLSEELTVALLFRKGRSDTAIKYEFISIPLEEEVDYM	344
Qy	596	-----KSSHFKDPRGSPVLSHNSVLSLSNQVEQPQDRP-----PAR	630
Db	345	YKFDNWFPCSYSCGKGVQTRNLYCIDCKNGRVEDDLCENNAKPEFEKSCETVDCEAE	404
Qy	631	WVAGSWGPCASCG-SGLQKRAVDC-RGSAGQRTVPACD---AAHRPVEVTAQGE-PCPT	684
Db	405	WFTGDWESCSSTCGDQGGQYRVVYCHQVFANGRRVTVEDGNCVTPRPVQTCNRFACPE	464
Qy	685	WELSAWPCSKSCGRGQRRLSKC---VHGGRLLARDOCNLHRKPQELD---FCVLRP	737
Db	465	WQAGPWSACSEKCGDAFYRSVTCRSEKEGEGKLLAADACPADEQ-EKFDTEKTCNLGP	523
Qy	738	C 738	
Db	524	C 524	
RESULT 7			
T34395			
hypoetical protein C37C3.6b - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000			
R:Geisel, C.; Bradshaw, H.			
C:Accession: T34395; T34394			
submitted to the EMBL Data Library, July 1996			
A:Description: The sequence of C. elegans cosmid C37C3.			
A:Reference number: Z21518			
A:Accession: T34395			
A>Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-2167 <GEI>			
A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GN00023; CESP:C37C3.6b			
A:Experimental source: strain Bristol N2; clone C37C3			
A:Accession: T34394			
A>Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1555, 'SKF' <GE2>			
A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a			
A:Experimental source: strain Bristol N2; clone C37C3			
C:Genetics:			
A:Gene: CESP:C37C3.6b; CESP:C37C3.6a			
A:Map position: 5			
A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 15			
Query Match 16.8%; Score 681; DB 2; Length 2167;			
Best Local Similarity 33.7%; Pred. No. 2e-38;			
Matches 162; Conservative 63; Mismatches 176; Indels 80; Gaps 19;			

[illegible]

Qy	387	LAVANWPKYSGVS-----PRDKCKLLCRANGTCYFYVLAPKVVVDGTLCSPOSTSVCVQGR	KCI	443
Db	534	-----LSGEGMQFQAAKCKTWLISJSTNIRTYS-NPFDGACGPG--QYCIKGE	CR	582
Qy	444	KAGCDGNLGSKKRFDKCGVCGDNKSKVKVTLFTKPMHGYNFVVAIPA	GASSIDIRQ	503
Db	583	PLLCGSTTAYSEAD-CPL-----SVLQTT--TTPPMHVHSV-----		617
Qy	504	YKGLIGDDNYLALKNSOCKY-----LLN---GHWFS-AVEROLVKGSLLRYS	G--TGTA	553
Db	618	-----DQFAGKTNPKYEHKHTPFLNEMSGWSVSECVTYDCHTGQVKVRVRCL	UAGVC	670
Qy	554	VESLQASRPFILEPLTVELVSWGKMTPPR--VRYSFYLPKEPRDKSSHPKPRG	PSVLHN	611
Db	671	AGALRERQPCTRPCTGSERL-TTSPPOQFNRNFIAPLNRQT-----N		714
Qy	612	SVLSLNOVQPPDDRPAPRVAVASGWCPCSASCSSGLQ-KRAVDCRGSAGORTVP	ACDAAH	670
Db	715	MIL-----RKVDHW--GPWSACSVTCGTGKLRRENCICQBCAETG	FCV----	757
Qy	671	RPVETQACGECPTW-ELSAWSPCSKSCGRGFORRLSK		708
Db	758	-----MOSCRENKNTWTWESOWSDCSVNCGRGVQFRK		792

RESULT 9

T18856
antagonensis inhibitor homolog - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18856; T24653
R:McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19031
A:Accession: T18856
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1444 <W1>
A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GNO0028; CESP:C02B4
A:Experimental source: clone C02B4

A: Monoclonal antibody; A:
submitted to the EMBL Data Library, July 1995
A: Reference number: Z19917
A: Accession: T24653
A: Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: DNA
A: Residues: 1-1444 <SWI>
A: Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1
A: Experimental source: clone T07C5
C: Genetics:
A: Gene: CESP:C02B4.1
A: Map position: X
A: Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3;

Query Match 12.5%; Score 504; DB 2; Length 1444;
Best Local Similarity 23.0%; Pred. No. 1.8e-26;
Matches 22; Conservative 91; Mismatches 306; Indels 348; Gaps 43;

QY 25 DLEHYL---LTLLATAARLYRHSILNPINIVKVKLLLRDRDS-----GPKVTGNAALT 76
 | | | | : | | | : | | | : | | | : | | | : | | | : | | | :
DB 239 DAEOHLLFEFLALNNVHVLIYOODITPNDLDIVIVRYEMWRTQPSALSTGVGHKNQAOSL 298
 | | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 77 LRNFCAHQKLNKYSD-KHEPEYWDTAILETRDQLCGACTCDTLGMADYGTMCDDPKRSCSV 135
 | | | | : | | | : | | | : | | | : | | | : | | | : | | | :
DB 299 LDAFCRYQAHNPGCTDLTDNNHYDHGVLLTGYDIYHTTT-SVAGVAPVARMCDPLFACSL 357
 | | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 136 IEDDGLLSAFPTTAHELGHVFNMHPDNVK-VCEEVEFKLRANHHMSPTLLIQIDRANPNASAC 194
 | | | | : | | | : | | | : | | | : | | | : | | | : | | | :
DB 358 VEGUHLGRSFVLAEHMGHNMGVHDGVQONCKGCCLMSAVNGACKT-----TWSDC 409
 | | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 195 SAAIIITDFL----DSHGHCDCLLDQPKPISLPE----DLFGASYTILSOCELAFIGVGSK- 245
 | | | | : | | | : | | | : | | | : | | | : | | | : | | | :
DB 410 SVRFENAFLLQDPSRGNCCLRASPCLISTNLHSDLRLPGORFTADOCYSFWGRDYKV 469
 | | | | : | | | : | | | : | | | : | | | : | | | : | | | :

F:46-106/Domain: thrombospondin type 1 repeat homology <THRL>

Query Match 12.2%; Score 493; DB 2; Length 951;
Best Local Similarity 22.2%; Pred. No. 6.3e-26;
Matches 171; Conservative 55; Mismatches 183; Indels 360; Gaps 27;

QY 305 GSWAKWDYGPGRSRTCGGQVQLARRQC-----TNPTPANGKYCEGVVRYKSCNLEPCPS 360
DB 51 GEMTKWTF--SRSCGGVTSQERHCLQORRKSVPGPGRNRTCTGTSKRYQLCRVQECF- 106
QY 361 SASGKSFRECEQEARNG--YHNHSTNRLTLAVAWPKYSCVSPRO-----KCKLICRAN 411
DB 107 -PDGRSFRDEQCVSNHYINGRTHQ-----WKPLY-----PDDYVHSSPPCDLHC--TT 154
QY 412 GTGYFVLAPKPVVDGTGCS-PDSTSVCVQKCIKAGCDGNLGSKKRFRKCGVCGDGNKSC 470
DB 155 VDGQQLMVP-ARDGTGSKLTDLRGVCGVSGKEPFGCDGVLFSTHTLDCGCGICQDGSNC 213
QY 471 KKVTLGFTK-PMH-GYNFVVAIPAGASSIDIRQYKGLIGDDNYLAKNSOGKYLNGH 528
DB 214 THVTGNRYKGNALHGYSLVTHIPAGARDIQIVERKKA-----DVLALADEAGYFFNGN 268
QY 529 FVVSAYERDLVVGSLRLRYSGT-----GTAVESLQASRPILPTVEVLSV-GKMTPPPRVR 583
DB 269 YKVDS-PKFNENIAGTVVYKRRPMDVYETGIEYIVAGQPTNOGLNVVWVWQNGK--SPSIT 325
QY 584 YSFYLPKEPREDKS-----KD-----PR----- 604
DB 326 FEYTLQPPHESRPPQIYYGFSAESQGLDAGLMGFIPHNGSLYGOASSERLGLDNRL 385
QY 598 -SHF----- 600
DB 386 FGHFGLDMELGSGQGTNEVCEQAGGACGEGPRGKGFDRNVTGTPLTGKDDDEVDY 445
QY 601 -----KD-----PR----- 604
DB 446 HFASQEFFSANALSDOLLGAGSDLDKFTLNETVNSIFAQAPRSSILAESFFVDYENEGA 505
QY 605 GPSVLHNSVLSL-----NOVEQPDPRPPA----- 629
DB 506 GPYLLNGSVLELSDRVANSSEAPFPNVSTSLTTSAGNRTHKARTPRKARKOGVSPADM 565
QY 630 -----R 630
DB 566 YRWKLSHPECSATCTGTWMSAYAMCVRDGVVEVDDSYCDALTRPEPVHFCAGRECQPR 625
QY 631 WVASGWSGPCSASCSGLQKRAV----- 652
DB 626 WETSSWSECSRTCGEGVQPRVRCWKMLSPGFDSSVYSDICEAAEAARVPERKTCRNPAC 685
QY 653 -----DCRGSAGQRTV-----PACDAARHPVETOAC-GEPCP-TWELSA 689
DB 686 GPQWEMSENESECTAKGERSVVTIRDCSEDEKLCDPNTRPVGEKNCCTGPPCDRQWTVSD 745
QY 690 WSPCKSKCGRGRFRRSLKCVGHGRLIARDQCNLHRKQPQLDFCFLRPC 738
DB 746 WGPCSGSGGGRTRIRHYCKTSDGRVVPESQCMETKPLAIHPGCGDKNC 794

RESULT 11
T15976
hypothetical protein F08C6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15976
R:Bentley, D.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F08C6.
A:Reference number: Z18440
A:Accession: T15976
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-957 <BEN>

QY 246 PCP-----YMQYCTKLWCTGAKGWMVCQTRHFPWADTSGEGKCLCKGACVE-RHNLNK 300
DB 470 EIPNGKAMDICRILWCGNSGS---TISTAH-PALEGSMCGANKWKHGQCTHTWFTGLTP 525
QY 301 HRVDGSAKWD-----PYGPCSRTCGGQVQLARRQCTNPTPANGKYCEGVVRYKSCN 354
DB 526 VPIDGENSEGGAARKGCPIOQCAVSGSITVQGHRCVNPAPNNGGKTCEGANIRGIVCG 585
QY 355 LEPCCPSASGSKFREEO-----CEAF-----NGYHNHSTNRLTLAVAWPKYS 396
DB 586 ---ATSSNCLGFTREFFGNKICSSIKYDPHKPDQQLTGEGFEHST----- 627
QY 397 GVSPRDCKLICRANGT-----GYFVVLAPKVVVDGTGLCSPDSTSVCVQKCIKAGCDGN 450
DB 628 ---QPCRVWCHLIGSELIRNKGOF-----PDGTFPGFD--AYCVGGQCLALSNDK 673
QY 451 ----- 450
DB 674 ALVEQPEDCPRIEGRSVHWEWSWSECSVSCGLGGRVREYRKCSSGRKCGVSESRP 733
QY 451 -----LGSKRF-----DKCG----- 461
DB 734 CEGVLRCDEEFGKENGWCSSEKCALGVQKFRPCLTDQCSSKHLQBERPCDNEGCVTNW 793
QY 462 -----VCGG-----DNKSKKVTGLFTKPMHGYNFVVAIPAGA 494
DB 794 DEWSSCSGSGGGRYRIKCLDKDGDGDLLEKESCN-----TOKCISOSGWDWLPSCV 847
QY 495 S---SIDIRQRYGKYLIGDDNYLALKNSQKYLNGHVFVSAVERDLVVGSLRLYS--- 548
DB 848 SCGIGFQIRER-----LCDGELCATANKQAR-TCNQOQCPSAF-----SLSVWSEWG 893
QY 549 -----GTGTAVESLQASRPILPTVEVLSVGVKMTPPRVYSF 586
DB 894 EWTTCSATCGEGLOSRSRSCRGCTEDDASQTRCVNGPCEHSHYLTWSEWTTTCETCSF 953
QY 587 YLPKE-----PREDKSSHPKDPGRGPSVLHNSVLSLSNOVEQ 622
DB 954 DSRKRIACDGTTCNDQKIDETCDIACLRKHSFGFISPRPKLI-----TSNDLRK 1007
QY 623 PDDRP-----PARWAGSWGSPCSASCSGLQKRAVDRCGSAGQRTVPACDAARHPE 674
DB 1008 AFRGRLPLPIESIHSEKW--SWMGPCSVTCGSGRRVTRGQCEA-----SCPEQH--IQ 1056
QY 675 TQACG-EPCTWEL---SAWSPCKSKCGR-GFQRRSLKCVGHGRLIARDQCNLH---RK 726
DB 1057 TEECNLNSC--LELFINDWSCKSKSCGQDGIQTRQKLC-----LFNNAECSSVYESRR 1108
QY 727 POELDFC 733
DB 1109 CKDLFSC 1115

RESULT 10
T00260
hypothetical protein KIAA0605 - human
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00260
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00260
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-951 <NAG>
A:Cross-references: EMBL:AB011177; NID:g3043733; PIDN:BA25531.1; PID:g3043734
A:Experimental source: brain
A:Genetics:
A:Note: KIAA0605
C:Superfamily: thrombospondin type 1 repeat homology

A:Cross-references: EMBL:U29378; NID:9868184; PID:9868185; PIDN:AA68721.1; CESP:F08C6.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F08C6.1
A:Introns: 23/1; 135/3; 220/3; 285/3; 325/2; 376/3; 461/2; 498/3; 532/3; 627/2; 714/1; 7

Query Match 10.2%; Score 412; DB 2; Length 957;
Best Local Similarity 22.8%; Pred. No. 2.3e-20;
Matches 186; Conservative 81; Mismatches 256; Indels 292; Gaps 38;

QY 8 VETLVVADESMVKF-----HG--AD--LEHYLLTLATAARLYRHPSILNPIVIVVVKLL 59
DB 221 VELAVEVDENLRHFSKGGMADRKLQDYTLTLNNIQIMYQYPTASPPTEFRVREV 280

QY 60 LRDRDSG---PKVTCNAALTRNCAWKKLNKYSKDHPEWDTAILFTQDL-CGATT 114
DB 281 LTRQPSALAGYLHNGNAQYLDRCRYQNL-AVRD-----WDHAIMLTGYDTHRGAGS 334

QY 115 CDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKCEEVFGKLRA 174
DB 335 RSISGARLDGMDPWNCTLAEGLDFTSAFIGTHELGHRFTELKDH-----381

QY 175 NHMWSPTLQIDRANPWSACSAIITDLDSDHGDCLLDQSPKPSLDPGLPGASYLSQ 234
DB 382 ----KSDTLF-----387

QY 235 QCELAFGVSKPCPYMOYCTKLWCTGKARGQWVCQTRHFPWADGTSCGEGKLCCLKGACVE 294
DB 388 -----TGTFGCPGSKWCQGRCPV 405

QY 295 RHNLNK-----HRVDGSAKWDPYGPQSR-TCGG-----GVQLARQCT 332
DB 406 WTCNTNIOPTQVHVAPVVTTLFSLRDSGWSGWA-TICSOCTCNGILSGVLAIRRTCS 464

QY 333 NPTPANGKYCEGVRYKRSNLEPCPSASGKSPREOC-----EAFNGYNHSTN 383
DB 465 APYPANGGDCVGTSTRAVLCSRQGRASKSVDEYISDKMEQKRLKNDRELTKGSQLN 524

QY 384 RLTLAVAWPKYSGVSPRDKCLKICRAN---GTGYFY-VLAPKVVDGTLCSPDSTSVCVQ 439
DB 525 RF-----PORACKVFCDVQOHYGSQRNRYFFGDLNPDGTSCTGYD-RYCLD 568

QY 440 GKCIKAGCDNL-----GSKKRDKC-GVCGGD---N 467
DB 569 GECLALNCNNALISRDQSCPTDCTPITDQSSVYRGOWGWSLWTSCTATCGGGYRKN 628

QY 468 KSCKVTGLFTKPMHGYNFVAIPAAGASSIDIRQRYKGLIGDDNYLALKNSQGYLLNG 527
DB 629 RAC-SITG-----QCEGNETEVCSSESCPSV-----LRVGNWSWTWEN 669

QY 528 HFVVSAYERDLVVKGSLLRYSGTGTAVESLQASRPILBP---LTVEVLSVGKMTPPRVY 584
DB 670 HCVSVC-----GRGSOARY-----RKCLSPHRTLAFDCPENKVT-NELRI 709

QY 585 SFYLPKPREDKSSHPDRGPSVLHNSVLSNQVEQDDRRP-----ARWAGSWG 638
DB 710 TFF-----KARSYIMCSVRCNKIKRNTISEKNIEVRSQDPCNAGVWGTW--GGWST 761

QY 639 CSASCSG--LQRAVD---CRGSAGQRTVPACDAHRPVTQAGEPCPT---WEL-SA 689
DB 762 CSTSCGPGTLVQRQTCNRPDCGSAHERR--SCNVA-----TCQNDGIWSLWNE 808

QY 690 WSPCKSGKGRGFORRLKCVGHG--GRLLARDQC 722
DB 809 WSDCSRCVCKGLRSRSCFCGSCMGASSEQOFCN 843

RESULT 12
T22545

hypothetical protein F53B6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22545

R:white, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19578
A:Accession: T22545
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1059 <WIL>
A:Cross-references: EMBL:Z81086; PIDN:CAB03121.1; GSPDB:GN00019; CESP:F53B6.2
A:Experimental source: clone F53B6
C:Genetics:
A:Gene: CESP:F53B6.2
A:Map position: 1
A:Introns: 38/3; 92/3; 131/3; 169/1; 236/2; 282/3; 349/2; 420/1; 576/3; 759/3;

Query Match 10.2%; Score 411.5; DB 2; Length 1059;
Best Local Similarity 25.7%; Pred. No. 2.8e-20;
Matches 112; Conservative 53; Mismatches 128; Indels 143; Gaps 16;

QY 306 SWAKNDPYGPCSRTCGGGVQLARQCTNPTPANGKYCEGVRYKRSNLEPCPSASGK 365
DB 25 SWAANSPWSSCTKTCGGGVSRQLRRCLT-----SKCSGESVRFKVCQAQKTCESKS--R 75

QY 366 SPREECEAFNGYNHSTNPLTLAVAWPKYSCVSPRDKCLKICRANGTYGYVLAPKVV 425
DB 76 LARDTICGG-----EIVSRGQCEVVCSRLLTGANFLM--RVDD 112

QY 426 GTLC-SPDSTSVCVGCKIKAGCDGNLGGKRFDCGVCGGDNKSKKVTGLFTKPMHY 484
DB 113 GTPCQAATSRACVSKGSCQIVGCDGLISSFFEDACGVCGGGDTFC-----158

QY 485 NFVVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQGYLL--LNGHFFVSAVERDLVVK 542
DB 159 -----DN-----GKFIWKVSEYITACASNCDDIVD- 183

QY 543 SULLRYSGTGTAVESLQASRPILPTVEVLSV-GKMTPPRVRYSYFLPKPREDKSSH 601
DB 184 ----WSGAGRSSTASIS-----QPIVVCVNAITGRVVEKICADKLRFK-----222

QY 602 DPRGPSVLHNSVLSNQVEQDDRRP-----PARWAGSWGPCSGSLQKRAVDCR 655
DB 223 -----VEARPCPMLICPSRWMAADWTECPVPHGCEKTRKREVCV 261

QY 656 GSAGQRTVPA---CDAHRPVTQAC-GEPCPTWELSNWSPCKSGKGRGFORSLKCVG 710
DB 262 QTAHNVTVHVDPDTCFNGTRPAAEENCVSTSCGRWEAGKWSKCTASCGGVRRRHVACVG 321

QY 711 -----HGRLLARDQC 721
DB 322 GSDCDEGGRPRQETTC 337

RESULT 13
S24789

Jararhagin C precursor - Jararaca (fragment)
N:Alternate names: single chain botrocetin
N:Contains: disintegrin-like 28K protein; hemorrhagic proteinase (EC 3.4.24.-)
C:Species: Bothrops jararaca (Jararaca)
C:Date: 20-Feb-1995 #sequence_revision 29-Aug-1997 #text_change 09-Jun-2000
C:Accession: S24789; J02245; A44463; A37958; J02373

R:Paine, M.J.I.
submitted to the EMBL Data Library, August 1992
A:Reference number: S24789
A:Accession: S24789
A:Molecule type: mRNA
A:Residues: 1-571 <PAI>
A:Cross-references: EMBL:X68251; NID:962467; PID:962468
R:Usami, Y.; Fujimura, Y.; Miura, S.; Shima, H.; Yoshida, E.; Yoshioka, A.; Hirano, I.
Biochem. Biophys. Res. Commun. 201, 331-339, 1994
A:Title: A 28 kDa-protein with disintegrin-like structure (jararhagin-C) purified from
A:Reference number: J02245; MUID:94256699; PMID:8198592
A:Accession: J02245
A:Molecule type: protein
A:Residues: 360-571 <USA>

A:Experimental source: venom
R:Palme, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.
J. Biol. Chem. 267, 22869-22876, 1992
A:Title: Purification, cloning, and molecular characterization of a high molecular weight
11y.
A:Reference number: A4463; MUID:93054601; PMID:1385408
A:Accession: A4463
A:Molecule type: mRNA
A:Residues: 1-23, 'Q', 25-92, 'G', 94-131, 'G', 133-169, 'Q', 171-571 <PA2>
A:CROSS-references: GB:X68251
A:Experimental source: venom gland
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIP:118104)
R:Fujimura, Y.; Titani, K.; Usami, Y.; Suzuki, M.; Oyama, R.; Matsui, T.; Fukui, H.; Sug
Biochemistry 30, 1957-1964, 1991
A:Title: Isolation and chemical characterization of two structurally and functionally di
A:Reference number: A37958; MUID:91129280; PMID:1993206
A:Accession: A37958
A:Molecule type: protein
A:Residues: 360-372, 'E', 374-378, 'X', 380-384 <FUJ>
A:Note: 361-Val was also found
C:Comment: Inhibits collagen- and ADP-induced platelet aggregation.
C:Superfamily: mouse meltrin alpha; disintegrin homology
C:Keywords: hydrolase; metalloproteinase; venom; zinc
F:360-571/Product: jararagin C #status experimental <MAT>
F:362-444/Domain: disintegrin homology <DIS>
F:295,299,305/Binding site: zinc (His) #status predicted
F:296/Active site: Glu #status predicted

Query Match 6.4%; Score 258.5; DB 2; Length 571;
Best Local Similarity 24.2%; Pred. No. 4.4e-10;
Matches 117; Conservative 62; Mismatches 183; Indels 121; Gaps 26;

QY 6 RVETLVVADESMVFHAGDLEHYLLTLATAARLYRHPSTLNPI-NIVVVKVLL--LRD 62
Db 159 KIIEFVVVDQTVTKNGDLD-----KIKARYELANIVNEIFRYLYMHVALGLEI 211
QY 63 RDSGPKVT--GNAULTLRNFCAMQKLNKVSXKHPYWDTAILFTRQDLGATTCDTLGM 120
Db 212 WSGDKITVKPDVDTLNSFAEWRTDLLTRKKH---DNAQLLTAIDFNG-----PTIGY 263
QY 121 ADVGIMCDPKRSCVIED---DGLPSAFTTAHELGHVFNPHDVKVCEEVFKLRANH- 176
Db 264 AYIGSMCHPKRSGVIGVDQVSPNLVAVIMAHMGHNIGIHDT-----GSCSCDGY 315
QY 177 --MMSPTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQP-----SKPISLPEDLPGA 228
Db 316 PCIMGPT-ISNPSKFFSCSYIQCWDFIMNHNPCEIINEPLGTDIISPPVCGNELN--- 371
QY 229 SYTLSSQCELFATGSGKPCPYQYCTKLWCTGKAKQWVCQTRHFPWADGTSCEGKLC 288
Db 372 --EVGEECD---CGTPENCONECCDA-----ATCKLK-----SGSQCHGD--- 407
QY 289 KGACVERHNLNKHRYD--GSAWKBDPYGCSRTPCGGVLARRQCTNPTPANGGKYCEGV 346
Db 408 ---CCQCKFSKSGTECRASMECDPAEHTCQSS-----ECPADVFKNGQPCLD- 455
QY 347 RVKYSRNLPCP-----SSAGSKSFREOCEAFNGYHNSTNRLTLAVAWY 392
Db 456 --NYGVCYNGNCPIYHQCYALFGADVAEEDSCFDKNGKGYGYCRKENGKKIPCA-- 511
QY 393 PKYSGVSPRD-KC-KLICRANGTG-----YFVVLAPK-----VVDGTLGSPDSTSV 440
Db 512 -----PEDVKGRGLYCKDNPSGQNNPKMFYSNDDEHKGMVLPETKCA--DGKVC 562
QY 441 KCI 443
Db 563 HCV 565

RESULT 14
T00026
brain-specific angiogenesis inhibitor 1 - human

N:Alternate names: BAIL protein
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 12-Feb-1999
C:Accession: T00026
R:Nishimori, H.; Shiratsuchi, T.; Urano, T.; Kimura, Y.; Kiyono, K.; Tatsumi, K.; Yos
submitted to the EMBL Data Library, June 1997
A:Reference number: Z14064
A:Accession: T00026
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1584 <NIS>
A:CROSS-references: EMBL:AB005297; NID:d1175078; PID:d1024528
A:Experimental source: brain
C:Genetics:
A:Gene: GDB:BAIL
A:CROSS-references: GDB:9838088; OMIM:602682
A:Map position: 8q24-8q24
C:Superfamily: thrombospondin type 1 repeat homology
F:408-462/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 6.3%; Score 255.5; DB 2; Length 1584;
Best Local Similarity 22.0%; Pred. No. 2.3e-09;
Matches 120; Conservative 42; Mismatches 174; Indels 209; Gaps 26;

QY 238 LAFGVGSKPCPYQYCTKLWCTGKAKQWVCQTRHFPWADGTSCEGKLC 291
Db 197 LAGRSRSHPCGIMQ--TPCACLGGEAGGP-----AAGPLAPRGDVLDRNAVAGGPE 245
QY 292 -CYERHNLNK--HRVDGSAKMDPYGCSRTPCGGVLARRQCTNPTPANGGKYCEGV 348
Db 246 NCLTSLTQDRGGHGTGKWLWGECTRDGGLQTRTCL-PAPGVEGGCGEVLE 304
QY 349 KYRSCNLEPC-----PSSASKSPRE-----BQCEAFNGYHNSTNRLTLAVAWPK 394
Db 305 EGRQCNREACPGAGRTSSRSQSLRSTARRREBELGDELQOF-GFPAPQGTGDPAAEEW--- 360
QY 395 YSGVSPRDKCLICRANGTGYFYVLAPKVVDGTLGSPDSTSVCGKCIKAGCDGNLGS 454
Db 361 ----SPWSVCSSTC---GEGW-----QTRTRFCVSSS-YSTQCSGGLREQ 397
QY 455 KRFDKCGVCGDGNKSKKVTGLTKPMHGYNFVVAIPAGASSIDIRQGYKGLIGDNYL 514
Db 398 RLCNNSAVC-----PVHG-----ANDEW- 415
QY 515 ALKNSQCKYLLNGHFVYSAVERDLVWKSLLRYSGTGTAVERLQASRPILPTVEVLSV 574
Db 416 -----SPWSLCSSTC 425
QY 575 GKMPPPRVRYFYLPKEPREDKSHPKDPRGSPVLSNLSNQVEQDDRP----- 627
Db 426 GRGFRDRTR-----TCRPPQFGG-----NPCSGPEKQTKFCNIAL 460
QY 628 -PARWVAG-----SWGPCASCSGSLQKRAVDCRGSAGQRTVPACDAHRPVTQACG- 679
Db 461 CPGRAVDGNWNEWSWSACSASCSSQGRQRTRECNBP-----SYGGAECQGHWWETRD 516
QY 680 EPCPT---WEL-SAWSPCSKSGRQFRRSLKCVG--HGGRL--ARD---QCNLHRKP 728
Db 517 QCCPVGDKWQAWAGWGSVCSTCGAGSORRERVCSPFFGGACGQPDQYRQCGTQRCPE 576
QY 729 ELDFC 733
Db 577 PHEIC 581

RESULT 15
JC5928
semaphorin F precursor - human
C:Species: Homo sapiens (man)
C:Date: 10-Apr-1998 #sequence_revision 08-May-1998 #text_change 17-Nov-2000
C:Accession: JC5928
R:Simmons, A.D.; Pueschel, A.W.; McPherson, J.D.; Overhauser, J.; Lovett, M.
Biochem. Biophys. Res. Commun. 242, 685-691, 1998

Search completed: April 29, 2003, 17:13:47
Job time : 34.07 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: April 29, 2003, 17:10:25 ; Search time 18.0541 Seconds
(without alignments)
2738.218 Million cell updates/sec
Title: US-10-009-332-1_COPY_213_583
Perfect score: 2042
Sequence: 1 FVSPRVETLVVADESNVK.....SCNLEPCPSASGKSPFREQ 371

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	A_Geneseq_101002.*			
	1:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*	950	22
	2:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*	967	19
	3:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*	967	20
	4:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*	968	22
	5:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*	967	20
	6:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*	950	21
	7:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*	837	20
	8:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*	837	22
	9:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*	625	23
	10:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*	905	22
	11:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*	680	21
	12:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*	890	20
	13:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*	890	22
	14:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*	889	22
	15:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*	870	21
	16:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*	930	20
	17:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*	929	21
	18:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*	947	22
	19:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*	1602	23
	20:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*	1629	23
	21:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*	1629	23
	22:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*	1916	23
	23:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*	1916	23

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2042	100.0	928	23	AAU72899
2	2042	100.0	950	23	AAU72899
3	2042	100.0	950	23	AAU72899
4	2042	100.0	952	23	AAU72899
5	1920.5	94.0	505	21	AAU72899
6	1767	86.5	381	21	AAU72899
7	1337	65.5	727	20	AAU72899
8	1337	65.5	896	21	AAU72899
9	1334	65.3	950	20	AAU72899
10	1334	65.3	950	22	AAU72899

11	1334	65.3	950	22	AAU72899	Human METH1. Homo
12	1334	65.3	967	19	AAU72899	Human integrin lig
13	1334	65.3	967	20	AAU72899	Human tango-71 pro
14	1334	65.3	968	22	AAU72899	Protein; SEQ ID 12
15	1321	64.7	967	20	AAU72899	Human secreted pro
16	1310.5	64.2	950	21	AAU72899	Amino acid sequenc
17	1182.5	57.9	837	20	AAU72899	Human aggrecanase
18	1182.5	57.9	837	22	AAU72899	Human aggrecanase
19	1182.5	57.9	840	21	AAU72899	Human aggrecanase
20	1181.5	57.9	837	21	AAU72899	Human aggrecanase
21	1181.5	57.9	837	22	AAU72899	Human aggrecanase
22	1181.5	57.9	837	22	AAU72899	Human aggrecanase
23	1153.5	56.5	625	23	AAU72899	Human aggrecanase
24	1128	55.2	905	22	AAU72899	Human aggrecanase
25	1127.5	55.2	680	21	AAU72899	Human aggrecanase
26	1126.5	55.2	890	20	AAU72899	Human aggrecanase
27	1126.5	55.2	890	22	AAU72899	Human aggrecanase
28	1123.5	55.0	889	22	AAU72899	Human aggrecanase
29	1112	54.5	930	20	AAU72899	Human aggrecanase
30	1110	54.4	929	21	AAU72899	Human aggrecanase
31	1105	54.1	929	21	AAU72899	Human aggrecanase
32	1100	53.9	870	21	AAU72899	Human aggrecanase
33	1046.5	51.2	947	22	AAU72899	Human aggrecanase
34	1046.5	51.2	1602	23	AAU72899	Human aggrecanase
35	1046.5	51.2	1629	23	AAU72899	Human aggrecanase
36	1046.5	51.2	1629	23	AAU72899	Human aggrecanase
37	1046.5	51.2	1916	23	AAU72899	Human aggrecanase
38	1046.5	51.2	1935	23	AAU72899	Human aggrecanase
39	1040.5	51.0	1073	21	AAU72899	Human aggrecanase
40	1030	50.4	1882	22	AAU72899	Human aggrecanase
41	1030	50.4	1934	22	AAU72899	Human aggrecanase
42	1019.5	49.9	874	22	AAU72899	Human aggrecanase
43	963.5	47.2	1907	23	AAU72899	Human aggrecanase
44	957.5	46.9	958	21	AAU72899	Human aggrecanase
45	950	46.5	1505	23	AAU72899	Human aggrecanase

ALIGNMENTS

RESULT 1	
AAU72899	AAU72899 standard; Protein; 928 AA.
ID	AAU72899 standard; Protein; 928 AA.
AC	AAU72899;
DT	26-FEB-2002 (first entry)
DE	Human metalloprotease partial protein sequence #11.
XX	Human; protease; PCR primer; cytostatic; immunomodulator; cardiant; vasotropic; antimitigraine; analgesic; endocrine; nootropic; tranquiliser; hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic; anorectic; antinflammatory; aspartyl protease; cysteine protease; metalloprotease; serine protease; cancer; haematopoietic; breast; colon; lung; prostate; cervical; brain; ovarian; bladder; kidney; pain; immune-related disease; cardiovascular disease; neuronal disease; migraine; sexual dysfunction; mood disorder; attention disorder; cognition disorder; hypotension; hypertension; psychotic disorder; dyskinesia; metabolic disorder; inflammatory disorder.
OS	Homo sapiens.
XX	WO200183782-A2.
PN	08-NOV-2001.
PD	04-MAY-2001; 2001WO-US14431.
PF	04-MAY-2000; 2000US-201879P.
XX	(SUG-) SUGEN INC.

PI Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 PI Payne V;
 XX WPI: 2002-041502/05.
 DR N-PSDB; AAS97182.
 XX Novel protease polypeptide useful for screening for substances that may
 PT be used to treat, e.g., cancers, immune-related diseases,
 PT cardiovascular disease, migraine, pain, psychotic and inflammatory
 PT disorders -
 XX
 PS Claim 28; Figure 2G; 232pp; English.
 XX The invention relates to an isolated, enriched, or purified protease
 CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
 CC screen for substances (S) that may modulate its activity. Administering
 CC S (which modulates protease activity in vitro) may be used to treat a
 CC disease or disorder selected from cancers (e.g., of tissues, of blood or
 CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
 CC brain, ovarian, bladder or kidney), immune-related diseases and
 CC disorders, cardiovascular disease, brain or neuronal-associated diseases
 CC (e.g., central or peripheral nervous system diseases, migraine, pain,
 CC sexual dysfunction, mood disorders, attention disorders, cognition
 CC disorders, hypotension, hypertension, psychotic disorders, neurological
 CC disorders and dyskinesias), metabolic disorders and inflammatory
 CC disorders. (I) may also be useful as a diagnostic tool for a disease or
 CC disorder such as those above. AAU72876-AAU72910 represent human
 CC protease amino acid sequences of the invention.
 XX
 SQ Sequence 928 AA;
 Query Match 100.0%; Score 2042; DB 23; Length 928;
 Best Local Similarity 100.0%; Pred. No. 1.2e-162;
 Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSILNINIVVVKVLL 60
 DB 238 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSILNINIVVVKVLL 297
 QY 61 RDRDSGPKVTGNAALTLRNFCWQKLNKVS DKHPYWDTAILETRQDLGGATTCTDGLM 120
 DB 298 RDRDSGPKVTGNAALTLRNFCWQKLNKVS DKHPYWDTAILETRQDLGGATTCTDGLM 357
 QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180
 DB 358 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 417
 QY 181 TLQIDRANPWSACSAIIITDLDSDHGDCLLDQSPKISLPEDLPASVYLSQQCELA 240
 DB 418 TLQIDRANPWSACSAIIITDLDSDHGDCLLDQSPKISLPEDLPASVYLSQQCELA 477
 QY 241 GVSGKPCPYMOYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKLCCLKGACVERHNLNK 300
 DB 478 GVSGKPCPYMOYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKLCCLKGACVERHNLNK 537
 QY 301 HRVDGSNAKWDYPGCSRTCCGGVQLARRQCTNPTPANGGKYCEGVKRYRSCNLEPCPS 360
 DB 538 HRVDGSNAKWDYPGCSRTCCGGVQLARRQCTNPTPANGGKYCEGVKRYRSCNLEPCPS 597
 QY 361 SASGKSFRREQ 371
 DB 598 SASGKSFRREQ 608
 RESULT 2
 AAG62299
 ID AAG62299 standard; protein; 950 AA.
 XX
 AC AAG62299;
 XX
 DT 23-AUG-2001 (first entry)
 XX Human metalloprotease MDTs6 protein.

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
 KW osteopathic; antiarthritic.
 XX Homo sapiens.
 OS
 XX WO200134785-A1.
 PN
 XX 17-MAY-2001.
 PD
 XX 10-NOV-2000; 2000WO-JP07917.
 PF
 XX 11-NOV-1999; 99JP-0321740.
 PR
 XX 16-MAY-2000; 2000JP-0144020.
 PR
 XX (YAMA) YAMANOUCHI PHARM CO LTD.
 PA (KAZU-) KAZUSA DNA RES INST.
 PA
 XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
 PI
 XX WPI: 2001-343602/36.
 DR N-PSDB; AAH41003.
 DR
 XX Metalloprotease with aggrecanase activity for treating joint diseases
 XX especially osteoarthritis
 PT
 XX Claim 1; Page 56-60; 85pp; Japanese.
 PS
 XX This invention relates to a metalloprotease with aggrecanase activity.
 CC The invention includes protein and DNA sequences of the metalloprotease,
 CC vectors containing the DNA, host cells transformed by the vectors, and
 CC antibodies directed against the metalloprotease. The antibodies, protein
 CC and DNA sequences can be used in the treatment and prevention of joint
 CC diseases, particularly osteoarthritis. The treatment may result in
 CC osteopathic and antiarthritic activity. The present sequence represents
 CC the metalloprotease of the invention termed MDTs6.
 XX
 SQ Sequence 950 AA;
 Query Match 100.0%; Score 2042; DB 22; Length 950;
 Best Local Similarity 100.0%; Pred. No. 1.2e-162;
 Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSILNINIVVVKVLL 60
 DB 213 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSILNINIVVVKVLL 272
 QY 61 RDRDSGPKVTGNAALTLRNFCWQKLNKVS DKHPYWDTAILETRQDLGGATTCTDGLM 120
 DB 273 RDRDSGPKVTGNAALTLRNFCWQKLNKVS DKHPYWDTAILETRQDLGGATTCTDGLM 332
 QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180
 DB 333 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 392
 QY 181 TLQIDRANPWSACSAIIITDLDSDHGDCLLDQSPKISLPEDLPASVYLSQQCELA 240
 DB 393 TLQIDRANPWSACSAIIITDLDSDHGDCLLDQSPKISLPEDLPASVYLSQQCELA 452
 QY 241 GVSGKPCPYMOYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKLCCLKGACVERHNLNK 300
 DB 453 GVSGKPCPYMOYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKLCCLKGACVERHNLNK 512
 QY 301 HRVDGSNAKWDYPGCSRTCCGGVQLARRQCTNPTPANGGKYCEGVKRYRSCNLEPCPS 360
 DB 513 HRVDGSNAKWDYPGCSRTCCGGVQLARRQCTNPTPANGGKYCEGVKRYRSCNLEPCPS 572
 QY 361 SASGKSFRREQ 371
 DB 573 SASGKSFRREQ 583
 RESULT 3

```
AAE22541
ID AAE22541 standard; Protein; 950 AA.
XX
AC AAE22541;
XX
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #2.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX
OS Homo sapiens.
XX
PN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Friddle CJ, Hilbun E;
XX
DR WPI: 2002-372123/40.
DR N-PSDB; AAD35569.
XX
XX
Novel nucleic acid encoding a human protease, useful as a hybridization
probe for screening libraries and assessing gene expression patterns -
Claim 6; Page 36-38; 41pp; English.
XX
The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridisation probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 950 AA;
Query Match 100.0%; Score 2042; DB 23; Length: 950;
Best Local Similarity 100.0%; Pred. No. 1-2e-162;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVSIPRYVETLVVADESWKVHGADLEHYLLTLLATAARLYRHPISILNPINIVVKVLL 60
DB 213 FVSIPRYVETLVVADESWKVHGADLEHYLLTLLATAARLYRHPISILNPINIVVKVLL 272
QY 61 RDRDSGPKVTGNAALTNRNFCWQKLNKVDKHPYWDTAIFLTRODLGATCTDLGM 120
DB 273 RDRDSGPKVTGNAALTNRNFCWQKLNKVDKHPYWDTAIFLTRODLGATCTDLGM 332
QY 121 ADVGTMCDPKRSCVIEDGLPSAFTTAHELGHVFNPHDNKVCVEVFGKLRANHMSP 180
DB 333 ADVGTMCDPKRSCVIEDGLPSAFTTAHELGHVFNPHDNKVCVEVFGKLRANHMSP 392
QY 181 TLIOIDRANPWSACSAIITFDLDSHGDCILLDDPSKPISLPDLPGASYTLSQOCELAF 240
DB 393 TLIOIDRANPWSACSAIITFDLDSHGDCILLDDPSKPISLPDLPGASYTLSQOCELAF 452
QY 241 GVGSKPCPYMQYCTKLWCTGKAKGMQVCQTRHFPWADGTSCEGKCLKGCACVERHNLNK 300
DB 453 GVGSKPCPYMQYCTKLWCTGKAKGMQVCQTRHFPWADGTSCEGKCLKGCACVERHNLNK 512
QY 301 HRVDGSKWAKNDPYGPCSRCTCGGVQLARRQCTNPTPANGKCYCEGVYKVRSCNLEPCPS 360
|||||
DB 513 HRVDGSKWAKNDPYGPCSRCTCGGVQLARRQCTNPTPANGKCYCEGVYKVRSCNLEPCPS 572
QY 361 SASGKSFREQ 371
|||||
DB 573 SASGKSFREQ 583
RESULT 4
AAU74751
ID AAU74751 standard; Protein; 952 AA.
XX
AC AAU74751;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-11 protein sequence.
XX
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis.
XX
OS Homo sapiens.
XX
PN WO200198468-A2.
XX
PD 27-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-US19178.
XX
PR 16-JUN-2000; 2000US-212336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
PR 14-JUL-2000; 2000US-216946P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Deleagane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Wallia NK, Yao WG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzal Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX
WPI: 2002-090437/12.
DR N-PSDB; ABK12894.
XX
XX
Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
in the diagnosis, treatment and prevention of gastrointestinal (e.g.
gastritis), cardiovascular (e.g. atherosclerosis) and cell
proliferative (e.g. cancer) disorders -
Claim 1; Page 144-146; 177pp; English.
XX
The present invention relates to twenty one new human proteases,
referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
polypeptides of the invention are useful in the diagnosis, treatment and
prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
myocardial infarction, autoimmune/inflammatory e.g. acquired
immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
proliferative e.g. cancer, developmental e.g. Duchenne and Becker
muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
epilepsy and Alzheimer's disease and reproductive e.g. infertility and
endometriosis disorders. Numerous other examples of each disorder are
given in the specification. The present protein sequence represents
the human protease PRTS-11 protein of the invention.
XX
SQ Sequence 952 AA;
Query Match 100.0%; Score 2042; DB 23; Length 952;
```


FT Misc-difference 308 /note= "xaa= any amino acid"
FT WO200053774-A2.
PN 14-SEP-2000.
XX 08-MAR-2000; 2000WO-US06227.
XX 08-MAR-1999; 99US-0264585.
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX Kelner GS, Clark M, Maki RA;
PI WPI; 2000-594326/56.
XX N-PSDB; AAA95831.
DR Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Claim 12; Fig 23; 129pp; English.
XX The present sequence is human metalloproteinase ADAMTS-5. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMTS. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.
XX Sequence 381 AA;
SQ Query Match 86.5%; Score 1767; DB 21; Length 381;
Best Local Similarity 93.3%; Pred. No. 4.6e-140;
Matches 322; Conservative 3; Mismatches 18; Indels 2; Gaps 2;
QY 28 HYLTLATARLHPSILNPINIVVKKVLLLRDRSGPKVTGNAALTNRNFCWOKKL 87
DB 2 HYRARAARAG-IFKHPSILNPINIVVKKVLLLRDRSGPKVTGNAALTNRNFCWOKKL 60
QY 88 NKVSDKHPEYWDTAILFTRODLGATTCDTLGMADVGTMCDDPKRSCSVIEDDGLPSAFTT 147
DB 61 NKVSDKHPEYWDTAILFTRODLGATTCDTLGMADVGTMCDDPKRSCSVIEDDGLPSAFTT 120
QY 148 AHELGHVFNPHDNNVKVEEVFGKLRANHMMSPTLIQIDRANPWSACSAAIITDLDLSDGH 207
DB 121 AHELGHVFNPHDNNVKVEEVFGKLRANHMMSPTLIQIDRANPWSACSAAIITDLDLSDGH 180
QY 208 GDCILDDQSKPISLPEDLPAGASYTILSQCELAFGVGKPCPYMOYCTKLMCTGKAKQMV 267
DB 181 GDCILDDQSKPISLPEDLPAGASYTILSQCELAFGVGKPCPYMOYCTKLMCTGKAKQMV 240
QY 268 CQTRHFWADCTSCGEGKCLKLGACVERHNLNKRVDSWAKWDPYPCSRCTCGGVOLA 327
DB 241 CQTRHFWADCTSCGEGKCLKLGACVERHNLNKRVDSWAKWDPYPCSRCTCGGVOLA 300
QY 328 RROCTNPTP-ANGGKCEGVKRVYRSCNLEPCPSSASGKSFREQ 371
DB 301 RROXHQXPPLTGKCEGVKRVYRSCNLEPCPSSASGKSFREQ 345
RESULT 7
AAW78435
ID AAW78435 standard; Protein; 727 AA.
XX
AC AAW78435;

XX 11-MAY-1999 (first entry)
XX Human ADAMTS-1 protein.
DE Metalloproteinase-disintegrin protein with a thrombospondin domain; ARDS;
XX ADAMTS-1; drug composition; foodstuff; leukocyte; thrombocyte; hepatitis;
KW blood; erythrocyte; inflammatory disease; rheumatoid arthritis; asthma;
KW nephritis; Crohn's disease; acute respiratory disease syndrome.
XX Homo sapiens.
OS WO9855643-A1.
XX 10-DEC-1998.
PD 03-JUN-1998; 98WO-JP02449.
PF 03-JUN-1997; 97JP-0160422.
PR (KURE) KUREHA CHEM IND CO LTD.
XX Hakozaaki M, Hirose K, Inoguchi E, Ishida Y, Ishioka K;
XX Kuno K, Matsushima K;
PI WPI; 1999-070277/06.
XX N-PSDB; AAX17990.
DR Human metalloproteinase-disintegrin protein with thrombospondin
XX domain - useful as leukocyte and thrombocyte decreasing and
PT erythrocyte increasing agent
PT Claim 1; Page 51-52; 82pp; Japanese.
XX This sequence represents a novel human metalloproteinase-disintegrin
CC protein with a thrombospondin domain (ADAMTS-1). The protein may be used
CC in drug compositions and foodstuffs, as an agent for decreasing the
CC leukocyte and thrombocyte blood count and increasing the erythrocyte
CC blood count, e.g. for treatment of inflammatory diseases such as
CC rheumatoid arthritis, hepatitis, nephritis, Crohn's disease, asthma
CC and ARDS.
XX SQ Sequence 727 AA;
Query Match 65.5%; Score 1337; DB 20; Length 727;
Best Local Similarity 63.5%; Pred. No. 1.3e-103;
Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;
QY 1 FVSPRYVETLVVADSVKVFHAGADLEHYLLTLATAARLYRHPSTILNPINIVVKKVLL 60
DB 13 FVSPRYVETLVVADSVKVFHAGADLEHYLLTLATAARLYRHPSTILNPINIVVKKVLL 72
QY 61 RDRSGPKVTGNAALTNRNFCWOKKLKVKSDKHPEYWDTAILFTRODLGATTCDTLGM 120
DB 73 HDQKGPVETVSNALTLRNFNCWQKQHPNPPDRDAEHYDTAILFTRODLGATTCDTLGM 132
QY 121 ADVGTMCDDPKRSCSVIEDDGLPSAFTTAEHLGHVFNPHDNNVKVEEVFGKLRANHMMS 180
DB 133 ADVGTMCDDPKRSCSVIEDDGLPSAFTTAEHLGHVFNPHDNNVKVEEVFGKLRANHMMS 192
QY 181 TLQIDRANPWSACSAAIITDLDLSDGHGDCILDDQSKPISLPEDLPAGASYTILSQCELA 240
DB 193 MSLNLDHSPWSPCSAYMITSLFDLNGHGECLMDKPNQIQLPGLDLPGLTLYDANRQCQFF 252
QY 241 GYGSKEPCP-YMOYCTKLMCTGKAKQMVQTRHFWADCTSCGEGKCLKLGACVERHNLN 299
DB 253 GEDSKHCPDAASTCTSLWCTGTSGGVLYVQCTKHFFWADCTSCGEGKCLKLGACVERHNLN 311
QY 300 KH---RVDGSWAKWDPYPCSRCTCGGVOLAARROCTNPTPANGGKCEGVKRVYRSCNLE 356
DB 312 KHEDTFHGSWGPWGPWGDGCSRTCGGVQYTMRECDNPVPKNGGKCEGVKRVYRSCNLE 371
QY 357 PCPSSASGKSFREQ 371

```

Db 372 DCPDN-NGKTFREEQ 385
      || : ||: |||||
RESULT 8
AAB21265
ID AAB21265 standard; Protein; 896 AA.
XX
AC AAB21265;
XX
DT 23-FEB-2001 (first entry)
DE
DE Mouse metalloproteinase ADAMTS-1.
XX
XX Mouse; ADAMTS-1; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; neurotropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytotatic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
OS Mus musculus.
XX
PN WO200053774-A2.
XX
PD 14-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US06237.
XX
PR 08-MAR-1999; 99US-0264585.
XX
PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
PI Kelner GS, Clark M, Maki RA;
XX
XX WPI; 2000-594326/56.
XX
DR Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Disclosure; Fig 17; 129pp; English.
XX
XX The present sequence is mouse metalloproteinase ADAMTS-1. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.
XX
XX Sequence 896 AA;
XX
XX Query Match 65.5%; Score 1337; DB 21; Length 896;
XX Best Local Similarity 63.2%; Pred. No. 1.7e-103;
XX Matches 237; Conservative 51; Mismatches 81; Indels 6; Gaps 4;
Oy 1 FVSPRYVETLVVADESVMKPFHGLADLEHYLLTLTAARLYRHPISILNPINVVVKVLL 60
Db 237 FVSPRYVETLVVADESVMKPFHGLADLEHYLLTLTAARLYRHPISILNPINVVVKVLL 296
Oy 61 RDRSGPKVTGNAALTLENFAWQKLNKYSDKHPEVWDYTAILETRDLCGATTCDFLGM 120
Db 237 YEEQKGPVETVNAALTLENFAWQKLNKYSDKHPEVWDYTAILETRDLCGATTCDFLGM 356
Oy 121 ADVGMPCPKSCSVIEDDGLPSAFTTAHELGHVFNPNHVKYCEVFGKLRANHWMP 180
Db 357 ADVGTVCDSRSCSVIEDDGLQAAFTTAHELGHVFNPNHVDKACASINGVTGDSHLMAS 416

```

```

Oy 181 TLIQIDRANPWSACSAIIITDFLDGHCCLLDQSPKIPISLPEDLPQASVTLSCQCELA 240
      | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 417 MLSSLDHSQWSPSCSAAYMTSFLDNGHGECLMDKPNQPIKLPDLPGTLYDANRQCQFT 476
Oy 241 GVGSKPCP-YMQYCTKLWCIGKAGQWVQTRHFPWADGTSCEGKLCGLKAGACVERHNLN 299
      | | | | | | | | | | | : | | | | | | | | | | | | | | | | | | |
Db 477 GEESKHCPDAASTCTTLWCTGTSGGLLVQCTKHFPWADGTSCEGKLCGLKAGACVERHNLN 535
Oy 300 KH--RVDGSAKWDPPYPCSRCTCGGGVQLARROCTNPTPANGKYGCEGVRYRSCNLE 356
      | | | | | | | | | | | : | | | | | | | | | | | | | | | | | | |
Db 536 KHFPATPVHSGWGPWPGDCSRTCGGGVQYTMRECDNPVPKNGSKYCEGKRVRYRSCNIE 595
Oy 357 PCPSASGKSFREEQ 371
      | | | | | | | | | | | : | | | | | | | | | | | | | | | | | | |
Db 596 DCPDN-NGKTFREEQ 609
      | | | | | | | | | | | : | | | | | | | | | | | | | | | | | | |
RESULT 9
AAY49501
ID AAY49501 standard; Protein; 950 AA.
XX
XX AAY49501;
AC AAY49501;
XX
DT 10-JAN-2000 (first entry)
DE
DE Human METH1 protein.
XX
XX Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
KW arterial-venous malformation; immune deficiency.
XX
XX Homo sapiens.
XX
PN WO9937660-A1.
XX
PD 29-JUL-1999.
XX
PF 22-JAN-1999; 99WO-US01313.
XX
PR 23-JAN-1998; 98US-0072298.
XX
PR 28-AUG-1998; 98US-0098539.
XX
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
XX
PI IrueLA-Arispe L, Hastings GA, Ruben SM;
XX
XX WPI; 1999-590684/50.
XX
XX N-PSDB; AAZ32000.
XX
XX New isolated metalloprotease thrombospondin polypeptides, useful for
XX treating hyperproliferative disorders, cancers or autoimmune disorders
XX
XX Claim 10; Fig 1; 457pp; English.
XX
XX AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
XX metalloprotease thrombospondin (METH) proteins METH1 and METH2
XX respectively. METH1 and METH2 have been found to be potent inhibitors of
XX angiogenesis both in vitro and in vivo. They can be used for treating
XX cancer and other disorders related to angiogenesis including abnormal
XX wound healing, inflammation, rheumatoid arthritis, psoriasis,
XX endometrial bleeding disorders, diabetic retinopathy, some forms of
XX macula degeneration, haemangiomas, and arterial-venous malformations.
XX They may be useful in treating deficiencies or disorders of the immune
XX system, by activating or inhibiting the proliferation, differentiation,
XX or mobilisation (chemotaxis) of immune cells. The etiology of these
XX immune deficiencies or disorders may be genetic, somatic, such as

```

XX
anti-arthritis therapeutic -
PT
a new matrix protease and its preparation for use as an anti-cancer agent

XX
DD
20 - NOV - 2000XX
DD
20 - NOV - 2000

XX PF 25-MAY-2000; 2000WO-US14462.
XX 25-MAY-1999; 99US-0318208.
PR 20-JUL-1999; 99US-0144882.
PR 10-AUG-1999; 99US-0147823.
PR 13-AUG-1999; 99US-0373658.
PR 22-DEC-1999; 99US-0171503.
PR 22-FEB-2000; 2000US-0183792.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
PA (JONA/) JONAK Z L.
PA (TRUL/) TRULLI S H.
PA (FORN/) FORNWALD J A.
PA (TERK/) TERRETT J A.
XX Iruela-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Fornwald JA, Terrett JA;
XX WPI: 2001-025136/03.
DR N-PSDB; AAC90057.
XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX Claim 15; Fig 1; 768pp; English.
XX The present sequence is human METH1 (ME for metalloprotease and TH for
CC thrombospondin). METH1 can be used for inhibiting angiogenesis in an
CC individual, and for treating cancer, benign tumours, an ocular angiogenic
CC disease, rheumatoid arthritis, psoriasis, delayed wound healing, nonunion
CC endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion
CC fractures, scleroderma, trachoma, vascular adhesions, myocardial
CC angiogenesis, coronary collaterals, cerebral collaterals, arteriovenous
CC malformations, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH1 can also be used in birth control. METH1 can also
CC be used in diagnostic methods for the prognosis of cancer.
XX SQ Sequence 950 AA;
Query Match 65.3%; Score 1334; DB 22; Length 950;
Best Local Similarity 63.5%; Pred. No. 3.2e-103;
Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;
QY 1 FVSTPRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSPILNPINIVVVKVLL 60
DB 236 FVSSHRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSPILNPINIVVVKVLL 295
QY 61 RDRDSGPKVTGNAALTRNFCAWQKLNKVS DKHPYWDTAIFLTRDLCGGATCTDGLM 120
DB 296 HDEOKGPEVTSNAALTRNFCAWQKLNKVS DKHPYWDTAIFLTRDLCGGATCTDGLM 355
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANHMSP 180
DB 356 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANHMSP 415
QY 181 TLTIQDRANPWSACSAIITDFLDSDHGDCLLDQPSKIPSLPDLPGASYTILSQOCELAF 240
DB 416 MSLNLDHSQWSPSCSAIYMTSFLDNGHGECMLDKPQNPQIPLGDLPTGSDYANRQOFTF 475
QY 241 GVSGKPCP-TWQYCTKLVCTGKAGQWVCOTRHPFHWADGTSCGEGKLCCLKGACVERHNLN 299
DB 476 GEDSKHCPDAASTCLWCTGTSGVLVCQTQKHPFHWADGTSCGEGKLCCLKGACVERHNLN 534
QY 300 KH---RVDGSGWAKWDYPGPCSRTCGGGVQLARRQCTNPTPANGGKYCEGVRVRYRSCNLE 596

DB 535 KHFDTPEHSGWGMGPWGDCSRTCGGGVQYTMRECDNPVKNGGKYCEGVRVRYRSCNLE 594
QY 357 PCPSSASGKSFREQ 371
DB 595 DCPDN-NGKTFREQ 608
RESULT 12
AAW80285
ID AAW80285 standard; Protein; 967 AA.
XX AC AAW80285;
XX 19-JAN-1999 (first entry)
XX Human integrin ligand polypeptide ITGL-TSP.
XX ITGL-TSP; integrin ligand; angiogenic disease; cancer; atherosclerosis;
XX chronic inflammatory disorder; rheumatoid arthritis; tissue remodeling;
XX macular degeneration; diabetic retinopathy; Alzheimer's disease; human;
XX restenosis.
XX Homo sapiens.
XX EP874050-A2.
XX 28-OCT-1998.
XX 27-JAN-1998; 98EP-0300575.
XX 24-APR-1997; 97US-0845496.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX Fronwald JA, Hastings GA, Jonak ZL, Terrett JA;
PI Trulli SH;
XX WPI: 1998-544643/47.
DR N-PSDB; AAV66508.
XX DNA encoding integrin ligand polypeptide ITGL-TSP - used to treat
PT angiogenic diseases; restenosis, Alzheimer's disease and in tissue
PT remodeling
XX Claim 11; Pages 6-9; 24pp; English.
XX This represents a human integrin ligand polypeptide ITGL-TSP. ITGL-TSP
CC polypeptides can be used in the treatment of angiogenic diseases such as
CC cancer, cancer metastasis, chronic inflammatory disorders, rheumatoid
CC arthritis, atherosclerosis, macular degeneration or diabetic retinopathy,
CC restenosis, Alzheimer's disease and tissue remodeling. They can be used
CC to treat a subject in need of enhanced activity or expression of the
CC ITGL-TSP polypeptide.
XX SQ Sequence 967 AA;
Query Match 65.3%; Score 1334; DB 19; Length 967;
Best Local Similarity 63.5%; Pred. No. 3.3e-103;
Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;
QY 1 FVSIERYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSPILNPINIVVVKVLL 60
DB 253 FVSSHRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSPILNPINIVVVKVLL 312
QY 61 RDRDSGPKVTGNAALTRNFCAWQKLNKVS DKHPYWDTAIFLTRDLCGGATCTDGLM 120
DB 313 HDEOKGPEVTSNAALTRNFCAWQKLNKVS DKHPYWDTAIFLTRDLCGGATCTDGLM 372
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANHMSP 180

Db 373 ADVGTVCDPSCSVIEDDGLQAFAFTTAHELGHVFNMPHDDAKQOCASLNGVNDSHMMAS 432
 QY 181 TLQIDRANPWSACSAIITDFDLSHGDCLLDQPSKIPISLPEDLPASVYTLSSQOCELAF 240
 Db 433 MLNLDHSDQSPWSPCSAYMITSFLDNGHGECLMDKPNQPIQLPGDLGTSYDANRQCQTF 492
 QY 241 GVSGKPCP-YMOYCTKLWCTGKAKGOMVCOTRHFPPWADGTSCGEGKCLGKACVERHNLN 299
 Db 493 GEDSKHCPDAASTCTLWCTGTSGGLVLCQTKHFPWADGTSCGEGKWCINCKVNTD-R 551
 QY 300 KH---RVDSWAKWDPYPCSRCTCGGVQLARRQCTNPTPANGKGYCEGVVRYRSCNLE 356
 Db 552 KHFDTPFHGSGWGMWGPWDCSRTCGGVQYTMRECDNVPVKNKGKCYCEGKVRYSNLE 611
 QY 357 PCPSSASGKSFREEQ 371
 Db 612 DCPDN-NGKTFREEQ 625

RESULT 13
 AAY04142
 ID AAY04142 standard; Protein; 967 AA.
 AC AAY04142;
 DT 15-JUN-1999 (first entry)
 XX Human Tango-71 protein.
 DE Human; Tango-71; Tango-73; Tango-74; Tango-76; Tango-83; diagnosis;
 KW detection.
 XX Homo sapiens.
 XX WO9907850-A1.
 XX 18-FEB-1999.
 XX 06-AUG-1998; 98WO-US16502.
 XX 05-SEP-1997; 97US-0058108.
 XX 06-AUG-1997; 97US-0054966.
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PI Goodearl ADJ, Holtzman DA;
 DR WPI; 1999-167426/14.
 DR N-PSDB; AAX19955.
 PT New TANGO polypeptides and nucleic acids encoding them - useful as
 PT diagnostic agents and for treating disorders caused by aberrant
 PT expression of TANGO
 XX Claim 8; Fig 1; 84pp; English.
 CC The present sequence represents human Tango-71. Tango polypeptides are
 CC useful for identifying compounds which bind the polypeptide via direct
 CC binding, competition binding assays or Tango-71, -73, -74, 76 or -83-
 CC mediated signal transduction. Tango polypeptides are also useful for
 CC identifying modulating compounds by determining effect on Tango activity.
 CC Tango polypeptides and nucleic acids are useful for diagnosing diseases
 CC related to aberrant expression of Tango, and Tango polypeptides are
 CC useful for raising antibodies which can be used in diagnostic assays for
 CC detection of Tango, and also for generating anti-idiotypic antibodies for
 CC prevention and protection.
 XX Sequence 967 AA;
 SQ

Query Match 65.3%; Score 1334; DB 20; Length 967;
 Best Local Similarity 63.5%; Pred. No. 3.3e-103;
 Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;

QY 1 FVSIPRIVETLVVADESMVXFHGADLEHYLLTLTAARLYRHPISILNPINIVVVKVLL 60
 Db 253 FVSSHRVETMLVADQSMAEFHGSLKHYLLTLFSAARLYKHPISIRNSVSLVVKILVI 312
 QY 61 RDRSDGPKVTGNAALTLRNFCAQKLNKYSDKHPEWDFAILFTRODLGGATTCDTLGM 120
 Db 313 HDQKGPVTSNAALTURNFCAQKLNKYSDKHPEWDFAILFTRODLGGATTCDTLGM 372
 QY 121 ADVGTMCDPKRSCSVIEDDGLQAFAFTTAHELGHVFNMPHDDAKQOCASLNGVNDSHMMAS 180
 Db 373 ADVGTMCDPKRSCSVIEDDGLQAFAFTTAHELGHVFNMPHDDAKQOCASLNGVNDSHMMAS 432
 QY 181 TLQIDRANPWSACSAIITDFDLSHGDCLLDQPSKIPISLPEDLPASVYTLSSQOCELAF 240
 Db 433 MLNLDHSDQSPWSPCSAYMITSFLDNGHGECLMDKPNQPIQLPGDLGTSYDANRQCQTF 492
 QY 241 GVSGKPCP-YMOYCTKLWCTGKAKGOMVCOTRHFPPWADGTSCGEGKCLGKACVERHNLN 299
 Db 493 GEDSKHCPDAASTCTLWCTGTSGGLVLCQTKHFPWADGTSCGEGKWCINCKVNTD-R 551
 QY 300 KH---RVDSWAKWDPYPCSRCTCGGVQLARRQCTNPTPANGKGYCEGVVRYRSCNLE 356
 Db 552 KHFDTPFHGSGWGMWGPWDCSRTCGGVQYTMRECDNVPVKNKGKCYCEGKVRYSNLE 611
 QY 357 PCPSSASGKSFREEQ 371
 Db 612 DCPDN-NGKTFREEQ 625

RESULT 14
 AAB50011
 ID AAB50011 standard; Protein; 968 AA.
 AC AAB50011;
 DT 19-MAR-2001 (first entry)
 XX Protein; SEQ ID 125.
 DE Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Webber syndrome;
 KW plaque neovascularisation; telangiectasia; haemophilic joint;
 KW angiofibroma; fibromuscular dysplasia; wound granulation;
 KW Crohn's disease; atherosclerosis; birth control.
 XX Homo sapiens.
 OS
 XX WO200071577-A1.
 XX 30-NOV-2000.
 XX 25-MAY-2000; 2000WO-US14462.
 XX 25-MAY-1999; 99US-0318208.
 XX 20-JUL-1999; 99US-0144882.
 XX 10-AUG-1999; 99US-0147823.
 XX 13-AUG-1999; 99US-0373658.
 XX 22-DEC-1999; 99US-0171503.
 XX 22-FEB-2000; 2000US-0183792.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (SMK-) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONK/) JONAK Z L.
 PA (TRUL/) TRULLI S H.

PA	(FORN/) FORNWALD J A.	XX	Human; secreted protein; fusion protein; gene therapy; protein therapy;
PA	(TERR/) TERRETT J A.	KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
XX		KW	developmental abnormality; foetal deficiency; blood; allergy; renal;
PI	Iruela-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;	KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
PI	Forwald JA, Terrett JA;	KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
XX		KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
DR	WPI; 2001-025136/03.	KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
XX		KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
PT	METH1 and METH2 polynucleotides and encoded polypeptides, used to	XX	
PT	inhibit angiogenesis in the treatment of disorders such as cancer,	OS	Homo sapiens.
PT	rheumatoid arthritis and psoriasis	XX	
XX		XX	
PS	Claim 15; Pages 759-763; 768pp; English.	PH	Key
XX		FT	Misc-difference 40
XX		FT	Misc-difference 45
CC	The present invention relates to human METH1 and METH2 (ME for	FT	Misc-difference 169
CC	metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).	FT	Misc-difference 293
CC	METH can be used for inhibiting angiogenesis in an individual, and for	FT	Misc-difference 297
CC	treating cancer, benign tumours, an ocular angiogenic disease,	FT	Misc-difference 557
CC	rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,	FT	/label= unknown
CC	vasculogenesis, granulations, hypertrophic scars, nonunion fractures,	FT	/label= unknown
CC	scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,	FT	/label= unknown
CC	coronary collaterals, cerebral collaterals, arteriovenous malformations,	FT	/label= unknown
CC	ischaemic limb angiogenesis, Osler-Webber syndrome, plaque	FT	/label= unknown
CC	neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,	FT	/label= unknown
CC	fibromuscular dysplasia, wound granulation, Crohn's disease or	FT	/label= unknown
CC	atherosclerosis. METH can also be used in birth control. METH can also	XX	
CC	be used in diagnostic methods for the prognosis of cancer. The present	PN	WO9856804-A1.
CC	sequence is a protein isolated in the present invention.	XX	
XX		PD	17-DEC-1998.
SQ	Sequence 968 AA;	XX	
	Query Match 65.3%; Score 1334; DB 22; Length 968;	XX	
	Best Local Similarity 63.5%; Pred. No. 3.3e-103;	XX	
	Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;	XX	
QY	1 FVSIPRYVETLVVADSWKFGHADLEHYLLTLLATAARLYRHPSTLNPINIVVKVLL 60	XX	
Db	254 FVSSHRYVETLVVADSWMAEFHSGSLKHLVLLTFLSVAARLYRHPSTLNPINIVVKVLL 313	XX	
QY	61 RDRDSGPKVTGNAALTRNFCANVKLNKVDKHPYWDTAILTRQDLGGATTCDTLGM 120	XX	
Db	314 HDEQKGEVTSNAALTRNFCANVKLNKVDKHPYWDTAILTRQDLGGATTCDTLGM 373	XX	
QY	121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDKVNKVEEVFGKLRANHMSP 180	XX	
Db	374 ADVGTVCDPSKCSVIEDDGLQAAFTTAHELGHVFNPHNDKVNKVEEVFGKLRANHMSP 433	XX	
QY	181 TLIDIDRANPWSACSAAILITDFLDSHGDCLLDQPSKIPSLPEDLPASVTLSSQCELEAF 240	XX	
Db	434 MSLNLDHSPWSPCSAYMITSFLDNGHGCEIMDKPQNPQLPDGTPSTIDANRQCQTF 493	XX	
QY	241 GVGSKPCP-YMQYCTKLWCTGKAGQWVCQTRHFPWADGTSCEGKLCCLKGACVERHNLN 299	XX	
Db	494 GEDSKHCPDAASTCSTLWCTGSGVLVCQTKHFPWADGTSCEGKWCINGKVCNKTD-R 552	XX	
QY	300 KH---RVDSGWAKWDYPGCSRTCGGQVQLARQCTNPTPANGKVCYGVRYVYRSCNLE 356	XX	
Db	553 KHFDTPFHSGMGMGPWGDCSRTCGGQVQYTMRECDNFPVKNKGKVCYGVRYVYRSCNLE 612	XX	
QY	357 PCPSASCKSFREQ 371	XX	
Db	613 DCPDN-NGKTFREQ 626	XX	
	RESULT 15	XX	(HUMA-) HUMAN GENOME SCI INC.
AAW78189		XX	Brewer LA, Ebner R, Ferrie AM, Feng P, Greene JM, Lafleur DW;
AC	AAW78189 standard; Protein; 967 AA.	PI	Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P;
XX		PI	Yu GL;
XX		XX	WPI; 1999-080881/07.
XX		XX	N-PSDB; AAX04374.
DT	13-APR-1999 (first entry)	XX	
XX		XX	New isolated human genes and the secreted polypeptides they encode -
DE	Human secreted protein encoded by gene 64 clone HOUQC17.	PT	useful for diagnosis and treatment of e.g. cancers, neurological
		PT	disorders, immune diseases, inflammation or blood disorders

XX Claim 11; Page 297-300; 380pp; English.

XX This sequence represents a secreted human protein encoded by the gene

CC clone detailed in the descriptor line. The gene can be used to generate

CC fusion proteins by linking to the gene to a human immunoglobulin Fc

CC portion (e.g. AAX04302) for increasing the stability of the fused

CC protein as compared to the human protein only.

CC The invention relates to 86 novel genes and their fragments (nucleic

CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)

CC which are useful for preventing, treating or ameliorating medical

CC conditions e.g. by protein or gene therapy. Also, pathological

CC conditions can be diagnosed by determining the amount of the new

CC polypeptides in a sample or by determining the presence of mutations in

CC the new polynucleotides. Specific uses are described for each of the 86

CC polynucleotides, based on which tissues they are most highly expressed in

CC (see AAX04311 for described uses).

XX

SQ Sequence 967 AA;

Query Match 64.7%; Score 1321; DB 20; Length 967;

Best Local Similarity 62.9%; Pred. No. 4e-102;

Matches 236; Conservative 50; Mismatches 83; Indels 6; Gaps 4;

QY 1 FVSIPRYVETLVADSMVKFHGADLEHYLLTLLATARLYRHPHSILNPINIVVVKVLL 60

DB ||| |||||:||||:||||:|||||:||||:||||:||||:||||:||||:||||: 60

253 FVSSHRYVETMLVADSMVAEFGSGLKHYLLTFLSVARLKKHFXIRNSVSLVVKILVI 312

QY 61 RDRDSGPKVTGNAALTNRNFCANQKKNKSDKHPEYWDTAILFTRODLGATTCDTLGM 120

DB ||| |||||:|||||:||||:||||:||||:||||:||||:||||:||||:||||: 120

313 HDEQKGPVETNAALTNRNFCNWKQHPNPPSDRDAEHYDTAILFTRODLGSGTCDTLGM 372

QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180

DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 180

373 ADVGTVCDPSRSCSVIEDDGLQAFTTAHELGHVFNPHDDAKQKQACSLNGVNDSHMAS 432

QY 181 TLIQIDRANPWSACSAIITDLSGHDCLLDQSPKIPISLPEDLPGASYTLQQCELAF 240

DB ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| 240

433 MSLNLDHSPWSPCSAYMITSLDNGHGECLMDKPNQPIQLPGDLPGTSYDANRQCQTF 492

QY 241 GVGSKPCP-YMQYCTKLWCTGKAGQMVQCTRHFPWADGTSCGGKGLCKLGACVERHNLN 299

DB || || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| 299

493 GEDSKHCPDAASTCTLWCTGTSGLVLCQTKHFPWADGTSCGGKGLCKLGACVERHNLN 551

QY 300 KH---RVDGSAKNDPYGCSRTCGGVQLARROCTNPTPANGKYCEGVKYSNLE 356

DB || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| 356

552 KHFTDPPHSGMGMGPGWDCSRTCGGVQVYTMREDNVPVKNGGKYCEGKRVYRSCNLE 611

QY 357 PCPSSAGKSFREQ 371

DB ||: |||:||||| 371

612 DCPDN-NGKTFREQ 625

Search completed: April 29, 2003, 17:20:31

Job time : 22.0541 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 6.01802 seconds
(without alignments)
1813.869 Million cell updates/sec

Title: US-10-009-332-l_COPY_213_583
Perfect score: 2042
Sequence: 1 FVSIPIRYVETLVVADESWMK.....SCNLEPCPSASGKSFREO 371

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2.6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/1/iaa/PCTRUS_COMB.pep.*
6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1334	65.3	967	4	US-09-130-491-2
2	1182.5	57.9	837	4	US-09-122-126B-2
3	1136	55.6	608	4	US-09-130-491-13
4	1128	55.2	905	4	US-09-369-364A-9
5	1112	54.5	930	4	US-09-369-364A-2
6	1110	54.4	930	4	US-09-122-126B-15
7	1030	50.4	1882	4	US-09-369-364A-13
8	1019.5	49.9	874	4	US-09-369-364A-15
9	814	39.3	997	4	US-09-369-364A-7
10	687	33.6	551	4	US-09-130-491-16
11	654	32.0	1224	4	US-09-930-872-4
12	646.5	31.7	245	4	US-09-369-364A-11
13	631.5	30.9	1081	4	US-09-369-364A-17
14	629	30.8	518	4	US-09-369-364A-22
15	625.5	30.6	1211	4	US-09-491-522-5
16	618.5	30.3	1205	4	US-09-491-522-11
17	603	29.5	859	4	US-09-369-364A-5
18	493.5	24.2	481	4	US-09-130-491-8
19	421.5	20.6	566	4	US-09-491-522-7
20	332	16.3	491	4	US-09-930-872-2
21	229.5	11.2	1170	1	US-08-313-288B-20
22	214	10.5	802	4	US-09-632-098-2
23	214	10.5	812	4	US-09-632-098-4
24	205.5	10.1	441	3	US-08-985-526-3
25	203.5	10.0	751	2	US-08-836-442-3
26	202.5	9.9	529	2	US-08-836-442-3
27	196.5	9.6	814	4	US-09-813-819-4

28	196.5	9.6	814	4	US-09-920-048-4	Sequence 4, Appli
29	196.5	9.6	855	4	US-09-813-819-2	Sequence 2, Appli
30	196.5	9.6	855	4	US-09-920-048-2	Sequence 2, Appli
31	193.5	9.5	239	5	PCT-US93-01652-1	Sequence 1, Appli
32	184	9.0	201	4	US-09-411-329C-1	Sequence 1, Appli
33	184	9.0	201	4	US-09-411-329C-1	Sequence 1, Appli
34	184	9.0	201	4	US-09-466-276-1	Sequence 1, Appli
35	184	9.0	201	4	US-09-411-329C-3	Sequence 3, Appli
36	184	9.0	462	4	US-09-411-329C-17	Sequence 17, Appli
37	183.5	9.0	202	4	US-09-411-329C-16	Sequence 16, Appli
38	183.5	9.0	203	4	US-09-411-329C-5	Sequence 5, Appli
39	183.5	9.0	203	4	US-09-411-329C-15	Sequence 15, Appli
40	183.5	9.0	203	4	US-09-411-329C-3	Sequence 3, Appli
41	183.5	9.0	203	4	US-09-466-276-3	Sequence 3, Appli
42	183.5	9.0	464	4	US-09-411-329C-14	Sequence 14, Appli
43	183	8.8	462	4	US-09-026-001A-16	Sequence 16, Appli
44	179	8.8	592	4	US-09-026-001A-14	Sequence 14, Appli
45	178.5	8.7	200	2	US-08-836-442-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; EARLIER FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-130-491-2

Query Match	65.3%	Score	1334	DB	4	Length	967
Best Local Similarity	63.5%	Pred. No.	6.2e-113				
Matches	238	Conservative	50	Mismatches	81	Indels	6
Gaps	4						
QY	1	FVSIPIRYVETLVVADESWMKFGADLEHYLLTLATAARLYRHPSTLNPINVVVKVLL	60				
Db	253	FVSSHRVETMLVADSMARFSGSLKHYLLTLFVAARLYRHPSTLNPINVVVKVLL	312				
QY	61	RDRSGPKVTGNAALTURNCAMQKLNKVDKHPYWDTAILFTRODLCGATTCDTLGM	120				
Db	313	HDEQKGPVTSNAALTURNCAMQKLNKVDKHPYWDTAILFTRODLCGATTCDTLGM	372				
QY	121	ADVTGMDPKRCSVIEDDGLPSAFTAHGLGHVFNPHDKNVCEVEYFGKLRAHNMSP	180				
Db	373	ADVTGMDPKRCSVIEDDGLPSAFTAHGLGHVFNPHDKNVCEVEYFGKLRAHNMSP	432				
QY	181	TLIOIDRANPWSACSAIITDLDLSDHGCCLLDQPSKIPSLPDLPGASYTLLSQCELA	240				
Db	433	MLSNLDHQPWSPCSAYMITSFLLNGHGECLMDKPNQIQLPGLPGTSDANRQCQTF	492				
QY	241	GVSGKPCP-YMYCTKLWCTGKAKGQWVCTRHFPWADTSCGEGKCLKACACVERHNL	299				
Db	493	GEDSKHCPDAASTCTLWCTGTSGGLVLCQTKHFPWADTSCGEGKCLKACACVERHNL	551				
QY	300	KH---RVDSWAKWDVPGCSRTCGGGVQLARQCTNPTPANGKYCEGVRYKYSCLNLE	356				
Db	552	KHFTDPPHSGWGMGPWGDCSRTCGGGVQVYTMRECDNPVKNKGKCEGVRYKYSCLNLE	611				

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-130-491-13

Query Match 55.6%; Score 1136; DB 4; Length 608;
Best Local Similarity 63.0%; Pred. No. 3.5e-95;
Matches 201; Conservative 39; Mismatches 73; Indels 6; Gaps 4;

QY 57 VLLLRDRSGPKVTGNAALTLRNFCAWOKLKNVSKDHPKHEWDTAILFTRODLGGATTCD 116
DB 4 ILVIHQKGEVTSNAALTLRNFCAWOKLKNVSKDHPKHEWDTAILFTRODLGGATTCD 63

QY 117 TLGMADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVFGKLRANH 176
DB 64 TLGMADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVFGKLRANH 123

QY 177 MMSPTLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKPISLPDLPGASVYLSQQC 236
DB 124 MMSPTLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKPISLPDLPGASVYLSQQC 183

QY 237 ELAFVGSGKPCP-YMQYCTKLWCTGKAGQWVCOTRHPFADGTSCGEGKCLKAGACVER 295
DB 184 QFTGEDSKHCPDAASTCTLWCTGSGVLCQTKHFPWADGTSCGEGKCLKAGACVER 243

QY 296 HNLNKH---RVDGSAWAKWDPYPCSRCTCGGVQVLRQCTNPTPANGGKYCEGVRYKYS 352
DB 244 TD-RKHFTDPFHSGWGMGMPMGDCSRCTCGGVQVLRQCTNPTPANGGKYCEGVRYKYS 302

QY 353 CNLEPCFSSASGKSFREQ 371
DB 303 CNLEPCFSSASGKSFREQ 320

RESULT 4
US-09-369-364A-9
; Sequence 9, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-8
US-09-369-364A-9

Query Match 55.2%; Score 1128; DB 4; Length 905;
Best Local Similarity 53.8%; Pred. No. 3e-94;
Matches 205; Conservative 64; Mismatches 100; Indels 12; Gaps 6;

QY 1 FVSIPIRYVETLVVADESVMKFKHGADEHYLLTLATAARLYRHPISILNPINIVVKKVLL 60
DB 229 FVSEARFVETLVVADASMAAFYGTDLQNHILTVMSAARIYKHPISIRNSVNLVVKVLIV 288

QY 61 RDRSDGPKVTGNAALTLRNFCAWOKLKNVSKDHPKHEWDTAILFTRODLGGATTCD 119
DB 289 EKERWGEVSDNGGLTLRNFCSQWRFRNPKPSDRHPEHYDTAILFTRODLGGATTCD 348

QY 120 MADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVFGKLRANHMS 179
DB 349 MADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVFGKLRANHMS 408

QY 180 PTLIQRANPWSACSAIITDFLDSHGDCLLDQPSKPISLPDLPGASVYLSQQC 237

QY 357 PCPSSASGKSFREQ 371
DB 612 DCPDN-NGKTFREQ 625

RESULT 2
US-09-122-126B-2
; Sequence 2, Application US/09122126B
; Patent No. 6451575
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-2

Query Match 57.9%; Score 1182.5; DB 4; Length 837;
Best Local Similarity 55.5%; Pred. No. 3.1e-99;
Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSIPIRYVETLVVADESVMKFKHGADEHYLLTLATAARLYRHPISILNPINIVVKKVLL 60
DB 213 FASLSREVFETLVVADOKMAAFHAGLKYLLTVMAAAAKAFKPSIRNPVSLVTVLVL 272

QY 61 RDRSDGPKVTGNAALTLRNFCAWOKLKNVSKDHPKHEWDTAILFTRODLGGATTCD 120
DB 273 GSGEGPQVGPSPAAQTLRSCFAWGRGLNTPEDSDPDHEDTALFTRODLGGATTCD 332

QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVFGKLRANHMS 179
DB 333 ADVGTMCDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVFGKLRANHMS 392

QY 180 PTLIQRANPWSACSAIITDFLDSHGDCLLDQPSKPISLPDLPGASVYLSQQCELA 239
DB 393 PVMAHVDPPEFSPCSARFTDNLGNGYCHCLDKPEAPLHLPVTPFGKDYDADROCLT 452

QY 240 FVGSGKPCPYM-QYCTKLWCTGKAGQWVCOTRHPFADGTSCGEGKCLKAGACVERHNL 298
DB 453 FGPDSRHCPLPPLPCCALWCGLHNGHAMCQTKHSPWADGTCPGPAQACMGRCGLHMDQL 512

QY 299 NKHRVD--GSAWAKWDPYPCSRCTCGGVQVLRQCTNPTPANGGKYCEGVRYKYSCLNLE 356
DB 513 QDFNIPQAGGWPMPGMPGDCSRCTCGGVQVLRQCTNPTPANGGKYCEGVRYKYSCLNLE 572

QY 357 PCPSSASGKSFREQ 371
DB 573 DCP-TGSALTREQ 586

RESULT 3
US-09-130-491-13
; Sequence 13, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16

```
Db 409 PFFTHVNTLPWSPCSAVYLTLLDGDHCLLDAPTSLVPLTGLPHSHSTYELDDQCK 468
QY 238 LAEGVSGKPCP---YMQYCTKLWCTGKAKGQWVCOTRH--FPWADGTSCEGKLCILKGAC 292
Db 469 QIFGPDFRHCNTSVEDICVQICARHRSDEPICHTKNGSLWADGTPCGPGHLCILDGSC 528
QY 293 VERHNLN--KHRVDGSAWMDPYGPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVYKY 350
Db 529 VLKEDVENKAVVDGDPWGPWQCSRTCGGGIQFSNRCDNMPQNGRFGCLGERVY 588
QY 351 RSCNLEPCSSASGKSFREQ 371
Db 589 QSCNTEECF--PNGKSFREQ 607

RESULT 5
US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

Query Match 54.5%; Score 1112; DB 4; Length 930;
Best Local Similarity 53.3%; Pred. No. 9e-93;
Matches 200; Conservative 59; Mismatches 106; Indels 10; Gaps 3;

QY 2 VSIPIRYVETLVWADSMVKFPGADLEHYLLTLTAARLYRHPSTILNPINIVVYKVLRL 61
Db 263 ISRAQVELLLVADSSMARLYGRGLQHYLLTANRLYSHASIHRLAVKVVVLT 322
QY 62 DRDGGPKVTGNAALTNRNCAWQKLNKYSDKHPEYWDTAILFTRODLCGATTCDTLGM 121
Db 323 DKDTSLEYSKNAATTLKNFCKWQHQNGLDDEEHYDAAILFTREDLCGHHSCTDLGMA 382
QY 122 DVGTMDPKRSCSVIEDDGLPSAFTAHGLGVNPHNDVNVKVEVFGKLRANHMSP 181
Db 383 DVGTCSPERSCAVIEDDGLHAFTVAHEIGHLLGLSHDDSKFCENFTEDKRLMSI 442
QY 182 LIQIDRANPWSACSAIITDLSHGDCCLLDQSPKISLPEDLPASYYLSQQCELAFG 241
Db 443 LTSIDASKPWSKTSATITEFLDDHGNCLLDLPRKILGPELPQGYVDATQOCNLTFG 502
QY 242 VGSKPCPYMOCYTKLWCTGKAKGQWVCOTRHPFPWADGTSCEGKLCILKGACVER---H 296
Db 503 PEYSVCGMDVCARLWCVAVRQGMVCLTKLPVAVGTPCGGRVCLQCKVCVDTKKYY 562
QY 297 NLNKHVDGSAWMDPYGPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVYKYRSCNLE 356
Db 563 STSSH--GNWGSWGPWGQCSRSCGGGVQFAYRHNCNPAPRNNGRYCTGKRAIYRSCSV 619
QY 357 PCPSSASGKSFREQ 371
Db 620 PCP--PNGKSFREQ 632

RESULT 6
US-09-122-126B-15
; Sequence 15, Application US/09122126B
; Patent No. 6451575
```

```
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 15
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-15

Query Match 54.4%; Score 1110; DB 4; Length 930;
Best Local Similarity 53.6%; Pred. No. 1.4e-92;
Matches 201; Conservative 56; Mismatches 108; Indels 10; Gaps 3;

QY 2 VSIPIRYVETLVWADSMVKFPGADLEHYLLTLTAARLYRHPSTILNPINIVVYKVLRL 61
Db 263 ISRAQVELLLVADSSMARLYGRGLQHYLLTANRLYSHASIHRLAVKVVVLT 322
QY 62 DRDGGPKVTGNAALTNRNCAWQKLNKYSDKHPEYWDTAILFTRODLCGATTCDTLGM 121
Db 323 DKDTSLEYSKNAATTLKNFCKWQHQNGLDDEEHYDAAILFTREDLCGHHSCTDLGMA 382
QY 122 DVGTMDPKRSCSVIEDDGLPSAFTAHGLGVNPHNDVNVKVEVFGKLRANHMSP 181
Db 383 DVGTCSPERSCAVIEDDGLHAFTVAHEIGHLLGLSHDDSKFCENFTEDKRLMSI 442
QY 182 LIQIDRANPWSACSAIITDLSHGDCCLLDQSPKISLPEDLPASYYLSQQCELAFG 241
Db 443 LTSIDASKPWSKTSATITEFLDDHGNCLLDLPRKILGPELPQGYVDATQOCNLTFG 502
QY 242 VGSKPCPYMOCYTKLWCTGKAKGQWVCOTRHPFPWADGTSCEGKLCILKGACVER---H 296
Db 503 PEYSVCGMDVCARLWCVAVRQGMVCLTKLPVAVGTPCGGRVCLQCKVCVDTKKYY 562
QY 297 NLNKHVDGSAWMDPYGPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVYKYRSCNLE 356
Db 563 STSSH--GNWGSWGPWGQCSRSCGGGVQFAYRHNCNPAPRNNGRYCTGKRAIYRSCSV 619
QY 357 PCPSSASGKSFREQ 371
Db 620 PCP--PNGKSFREQ 632

RESULT 7
US-09-369-364A-13
; Sequence 13, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1882
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-9
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (468)
; OTHER INFORMATION: Xaa = C
; NAME/KEY: MOD_RES
; LOCATION: (521)
; OTHER INFORMATION: Xaa = Y
US-09-369-364A-13
```

Query Match 50.4%; Score 1030; DB 4; Length 1882;
Best Local Similarity 51.3%; Pred. No. 6.2e-85;
Matches 193; Conservative 58; Mismatches 109; Indels 16; Gaps 8;

QY 1 FVSIPRYVETLVADSWKFKGADLEHYLLTLATLAARLYRHPSILNPINIVVKKVLL 60
DB 236 FLSYPRFVEVLVADNRVSYHGENLQHYLLTMSIVASYKDSIGNLINIVINLVI 295
QY 61 RDRDSGPKVTGNAALTLRNFCAMQKLNKVS DKHPEYWDTAILFTRODLGCA--TTCDTL 119
DB 296 HNEQDPSISFNAQTLLKNFCQWHSNPGGIH---DTAVLLTRQDICRAHDKCDTL 351
QY 120 MADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANHMS 179
DB 352 LAELGTICDPYRSCSISDGLSTAFTHAHELGHVFNPHDNKNCKEE--GVKSPQHVA 410
QY 180 PTLIQIDRANPWSACSAIITDFLDSHGDCLLDQ--SKPISLPEDLPFGASYTILSQOCEL 238
DB 411 PTLNFTNPMWMSKRSKYIETFLDTGYGECLLNEPESRPYPLPVQLPGILYNVKNQXEL 470
QY 239 AFGVSGKPCPYMYCTKLWC---TKAKGQMVQCTRHPPWADGTSCGEGKLCGLKACVER 295
DB 471 IFPGSQVCPYMQCRRLWCNNVNGVHKG---CRTQHTPWADGTCEPEGKCKXGFCVPK 527
QY 296 HNLNKHRYDGSWAKWDPYPCSRCTCGGVQVLARROCTNPTPANGGKYCEGVRYKRSCL 355
DB 528 -EMDVPVTDGWSGWSPTGTCRTCGGKIKAIRECNPPEKNGKYCVGRMKFKSCN 586
QY 356 EPCPSSASGKSPREQ 371
DB 587 EPCLKQK--RDFRDEQ 600

RESULT 8
US-09-369-364A-15
; Sequence 15, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 874
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-9
US-09-369-364A-15

Query Match 49.9%; Score 1019.5; DB 4; Length 874;
Best Local Similarity 50.9%; Pred. No. 2.1e-84;
Matches 192; Conservative 59; Mismatches 109; Indels 17; Gaps 9;

QY 1 FVSIPRYVETLVADSWKFKGADLEHYLLTLATLAARLYRHPSILNPINIVVKKVLL 60
DB 127 FLSYPRFVEVLVADNRVSYHGENLQHYLLTMSIVASYKDSIGNLINIVINLVI 186
QY 61 RDRDSGPKVTGNAALTLRNFCAMQKLNKVS DKHPEYWDTAILFTRODLGCA--TTCDTL 118
DB 187 HNEQDPSISFNAQTLLKNFCQWHSNPGGIH---DTAVLLTRQDICRAHDKCDTL 242
QY 119 GMADVGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANHMM 178
DB 243 GLAELGTICDPYRSCSISDGLSTAFTHAHELGHVFNPHDNKNCKEE--GVKSPQHVA 301
QY 179 SPTLIQIDRANPWSACSAIITDFLDSHGDCLLDQ--SKPISLPEDLPFGASYTILSQOCE 237
DB 302 APTLNFTNPMWMSKRSKYIETFLDTGYGECLLNEPESRTPYPLPSQLPGLLYNVKNQCE 361

Query Match 39.9%; Score 814; DB 4; Length 997;
Best Local Similarity 46.8%; Pred. No. 1.2e-65;
Matches 176; Conservative 43; Mismatches 141; Indels 16; Gaps 10;

QY 2 VSIPRYVETLVADSWKFKGADLEHYLLTLATLAARLYRHPSILNPINIVVKKVLL 60
DB 238 VSKKWCETLVADNRVSYHGENLQHYLLTMSIVASYKDSIGNLINIVINLVI 297
QY 61 RDRDSGPKVTGNAALTLRNFCAMQKLNKVS DKHPEYWDTAILFTRODLGCA--TTCDTL 118
DB 298 EDEEDLKITHADNTLKSCKWKSINMKGDAPHLHDTAILLTRKDLCAAMNRPCE 357
QY 119 GMADVGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANHMM 177
DB 358 GLSHVAGMCQPHRSCSINEDTGLPFAFTVAHELGHSGIHDGSGNDCEPV--GK--RPFI 414
QY 178 MSPTLIQIDRANPWSACSAIITDFLDSHGDCLLDQSPKP--ISLPEDLPFGASYTILSQOC 236
DB 415 MSPQLLYDAAPLTWSRCSRYITRFLDRGWLCLDDPPAKDIIDFPSPVPGVLYDVSHQC 474
QY 237 ELAFGVSGKPCPYM--QYCTKLWCCTGKAKGQMVQCTRHPPWADGTSCGEGKLCGLKACVER 295
DB 475 RLQYGAISAFCEMDMNVTCLWCSVGT---TCHSKLDAADVDTGRCGENKWLCSGECVP-- 529
QY 296 HNLNKHRYDGSWAKWDPYPCSRCTCGGVQVLARROCTNPTPANGGKYCEGVRYKRSCL 355
DB 530 VGFREPAVDGWSGWSAWSICSRSCGMGVQSAERQCTQPTPKYKGYCVGERKFRCLNL 589
QY 356 EPCPSSASGKSPREQ 371
DB 590 QACP--AGRPSPRHVQ 603

RESULT 10
US-09-130-491-16
; Sequence 16, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.

; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-130-491-16

Query Match 33.6%; Score 687; DB 4; Length 551;
Best Local Similarity 56.4%; Pred. No. 2e-54;
Matches 119; Conservative 29; Mismatches 57; Indels 6; Gaps 4;
QY 165 CEEVFKLRANHMSTLQIDRANPWSACSAIITDFLDGSHGCLDQPSKPISLPED 224
Db 1 CASLNGVSGDLSHLSMLSLDHSOPSPCSAYMVTSLDNGHGECLMDKPNKLPSPD 60
QY 225 LPGAAYTLQOCELAFVGSKPCP-YMOYCTKLWCTGKAKGOMVQOTRHPFWADTSCGE 283
Db 61 LFGTLNANRQCOFTFGEESKUCPDAASTCTTLWCTGSGGLLVQOTKHPFWADTSCGE 120
QY 284 GKLCAGACVERHNKHK---RVDSWAKWDYPGCSRCTCGGGVQLARRQCTNPTTPANGG 340
Db 121 GKWCVSGKCVNKTDK-KHFAFVHSGWGPWGDGCSRCTCGGGVQYTMRECDNPVPKNGG 179
QY 341 KYCEGVRYKRCNLEPCPSSASGSKSFREQ 371
Db 180 KYCEGKRVRYRSCNIEDCDPDN-NGKTFREQ 209

RESULT 11
US-09-930-872-4
; Sequence 4, Application US/09930872
; Patent No. 6448388
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. 6448388 Human Proteases and Polynucleotides Encoding the Sa
; FILE REFERENCE: LEX-0219-USA
; CURRENT APPLICATION NUMBER: US/09/930,872
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,852
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1224
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-930-872-4

Query Match 32.0%; Score 654; DB 4; Length 1224;
Best Local Similarity 39.5%; Pred. No. 5.6e-51;
Matches 149; Conservative 49; Mismatches 149; Indels 30; Gaps 12;
QY 8 VETLVVADESVMKFKHG-ADLEHYLLTLATAARLYRHPHSILNPINIVVVKVLLLRDRDSG 66
Db 292 VETLVVDDKMMQNHGHNITVYILNMVSAFLKDGDTIGGNINIAIVGLILLEDEQPG 351
QY 67 PKVTGNAALTNRNCAWKKL-NKVSCKHPEYWDITLFTRODLGG--ATTCDTLGWADV 123
Db 352 LVISHHADHTLSSFCQWQSLMGKDGTRH---DHAULTGLDICSWKNEPCDTLGFAP 407
QY 124 GTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEVFVKLRANHMSP 182

Db 408 SGMCKYRSCNTINEDTGLGLAFTTAHESGHNFHMGEGNMCKKSEG-----NIMSP 462
QY 183 IQIDRANPWSACSAIITDFLDGSHGCLDQPSKPI---SLPEDLPASVTLTQQCELA 239
Db 463 AGRNGVFSWSPCSQYLLHKLFLSTAQAICLADQP-KPVKEYKYPEKLPGLYDANTQCKWQ 521
QY 240 FGVGSKPCPY---MOYCTKLWC---TGRKAGOMVQOTRHPFWADTSCGEGKLCUKGACVE 294
Db 522 FGEKALCMLDFKKDICKALWCHRIGRK-----CETKFMFAAEGTICGHDWMCRRGGCVK 576
QY 295 RHNLNKHRVDGWSWAKWDYPGCSRCTCGGGVQLARRQCTNPTTPANGGYKCEGVRYKYS 354
Db 577 YGDEGPKPTHGWDSDWSWSPCSRTCTCGGGVSHRSLCTNPAKPSHGGKFCGSTRTLK 636
QY 355 LEPDPSSASGSKSFREQ 371
Db 637 SOKCPRDS--VDFRAAQ 651
RESULT 12
US-09-369-364A-11
; Sequence 11, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-8
US-09-369-364A-11

Query Match 31.7%; Score 646.5; DB 4; Length 245;
Best Local Similarity 57.9%; Pred. No. 3.5e-51;
Matches 114; Conservative 41; Mismatches 41; Indels 1; Gaps 1;
QY 1 FVSIPRVETLVVADESVMKFKHGADLEHYLLTLATAARLYRHPHSILNPINIVVVKVLL 60
Db 21 FVSEARFETLVVADASMAAFYAGDLQNLHILTLMSVAARIYKHPKSIKNSLNMVVKVLI 80
QY 61 RDRSGPKVTGNAALTNRNCAWKKLNVKYSCKHPEYWDITLFTRODLGGAT-TCDTL 119
Db 81 EDEKGVSDNGGLTLRNFCNWRNQPSPDRHPEHYDTAILLTRNFCGQGLCDTLG 140
QY 120 MADVGTMCDPKRCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFVKLRANHM 179
Db 141 VADGTICDPKSCSVIEDGELQRAHTLAHELGHVLSNPHDDSKPCTRLFGPMKHHVMA 200
QY 180 PTLIQIDRANPWSACSA 196
Db 201 PLFVHLNQTLPWSPCSA 217
RESULT 13
US-09-369-364A-17
; Sequence 17, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-10
US-09-369-364A-17

Query Match          30.9%; Score 631.5; DB 4; Length 1081;
Best Local Similarity 37.1%; Pred. No. 5.3e-49;
Matches 145; Conservative 57; Mismatches 154; Indels 35; Gaps 12;

QY 2 VSPRIVETILVADENKVFHG-ADLEHYLLTLATAARLYRHPSTILNPINIVVVKVLL 60
DB 213 VSRERIVETMDVADKMMVAYHGRDVEQVVLAIMNIVAKLFQDSSLGSTVNILVRLILL 272
QY 61 RDRSDGPKVTGNALILRNFCAWQKL-----NKVSKHPEYWDTAILFTRODLC--G 111
DB 273 TEDQPTLEITHHAGSLDSFCWKQKSIVNHSHGNAIPENGVANHDTAVLITRYDICIYK 332
QY 112 ATTCDTILGMADVCTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEVF- 170
DB 333 NKPCGTLGLAR-WAECVSAAREAAASRTLAATSVHHCHEIGHTFGNHDGVGNSCGARQ 391
QY 171 ---KLNRANHMSPTLIQIDRANP--WSACSAIITDPLDSGHGCDLLDQPSK-PISLPED 224
DB 392 DPAKLMAAHITMKT-----NPFVSSCNRDYITSLDGLCLNLRPPRODFVYPTV 444
QY 225 LPGAASYTISOCELAFGVGSKPCPYWQYCTKLWCTGKAKGOMVQOTRHPFWADGTSCGEG 284
DB 445 APGQAYDADEQCFQHGKVKSRQCKYGEVSELWCLSKSNR---CITNSIPAAEGTLCQTH 501
QY 285 KL-----CLKGACVERHNLKHRVDGSAKWDPYGPCSRTCGGVQGLARRQCTNPTPANG 340
DB 502 TIDKGCWYKRVCP-FGSRPEGVDGANGPTWPGDCSRTCGGVSSSRHCDSPRTIGG 560
QY 341 KYCEGVKRVKRSNLEPCPSASGKSPREQ 371
DB 561 KYCLGERRRHSRSCNTDDCPGGS--QDPREVQ 589

RESULT 14
US-09-369-364A-22
; Sequence 22, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-5
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (99)
; OTHER INFORMATION: Xaa = M
US-09-369-364A-22

Query Match          30.8%; Score 629; DB 4; Length 518;
Best Local Similarity 48.9%; Pred. No. 3.5e-49;
Matches 110; Conservative 31; Mismatches 74; Indels 10; Gaps 3;

QY 152 GHVFNNPHDNVVKCEVFGLRANHMSPTLIQIDRANPWSACSAIITDPLDSGHGDC 211
DB 1 CHLLGLSHDDSKFCETFGSTEDKRLMSILTSIDASKPWSKTSATITEPLDDGHGNC 60
QY 212 LDQPSKPSLPEDLPGASYTISOCELAFGVGSKPCPYWQYCTKLWCTGKAKGOMVQOTR 271
```

```
DB 61 LDLPKQILGPEELPGQTYDATQCNTLTFGPDYSVCXGVCAKLMCAVVRQSGMCLTK 120
QY 272 HFPWADGTSCGEGKCLKGACVER-----HNLNKHRYDVGSAKWDPYGPCSRTCGGVQ 326
DB 121 KLPAVEGTPCGKGRICLOGKVDTKKKYYSTSSH---GNWGSWGSWGQCSRSRCSGGGVQ 177
QY 327 ARROCTNPTPANGKYCEGVKRVKRSNLEPCPSASGKSPREQ 371
DB 178 AYRHCNPNAPRNGRYCTGKRAIVHSCSLMPCP--PNGKSFRRHQ 220

RESULT 15
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lapiere, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennle & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING AGENT:
; ATTORNEY/AGENT INFORMATION:
; NAME: Adams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-5

Query Match          30.6%; Score 625.5; DB 4; Length 1211;
Best Local Similarity 38.5%; Pred. No. 2.1e-48;
Matches 145; Conservative 54; Mismatches 141; Indels 37; Gaps 15;

QY 11 LVVADESMVKFHGAD-LEHYLLTLATAARLYRHPSTILNPINIVVVKVLLLRDRSGPKV 69
DB 271 LLGVDDSVVQFHGKEHVQKYLTLTMNIVNEIYHDESLGAHINVLVRLILLSYGKMSLI 330
QY 70 -TGNAAALTNRNCAWKLNKLVSDKHPEYWDTAILFTRODLCGATTCDTLGMADVGTMD 128
DB 331 EIGNPQSLENVCRWAYLQOKPDTGDEYHDHAIFLTRQDF-GPSGMQ--GVAPVTGMCH 387
QY 129 PKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEVFGKLRANHMSPTLIQIDR 187
DB 388 PVRSCITLNHEDGSSAFVVAHETGHLGMEHDGQNRCDG---EVRGLSIMAPLVQAAPH 444
```

```

QY 188 ANPWSACSAIITDFLDGHDCLLDOP---SKPISLPEDLPASGYTLSSQOCELAFGVGS 244
Db 445 RFHWSRCSQOELSRYLHS--YDCLDDPPAHDWP-ALPO-LFGLHYSMNEQCRFDGGLGY 500
QY 245 KPCPYMXY---CTKLWCTGKAKGOMVCQTRHFFPWADGTSCGEGKLCCLKGACV-----E 294
Db 501 MMCTAFRTFDPCKQLWCS-HPDNPYFCYKKGPPLDGTMCAFGKHCFCGHCICWLTPTDILK 559
QY 295 RHNLNKHRYDGSWAKWDYPGCSRTCGGVQVLARQCTNPTPANGGKYCEGVVRKYRSCN 354
Db 560 R-----DGSWGAWSPEGSCSRTCGTGKTRQCONPHANGRTCSGLAYDFQLCS 611
QY 355 LEPCPSSASGKSFRREQ 371
Db 612 RQDCPDSLA--DFREQ 626

```

Search completed: April 29, 2003, 17:11:47
Job time : 11.018 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:12:21 ; Search time 8.24692 Seconds
(without alignments)
3604.758 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_583

Perfect score: 2042

Sequence: 1 FVSIPRYVETLVVADESWK.....SCNLEPCPSASGKSFRESEQ 371

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pap:*
- 2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pap:*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pap:*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pap:*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pap:*
- 6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pap:*
- 7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pap:*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pap:*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pap:*
- 10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pap:*
- 11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pap:*
- 12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pap:*
- 13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pap:*
- 14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2042	100.0	950	10	US-09-965-631-4
2	2020.5	98.9	823	9	US-10-163-316-2
3	1339	65.6	727	9	US-10-097-597-12
4	1339	65.6	727	9	US-10-097-580-12
5	1339	65.6	727	10	US-09-445-023A-12
6	1337	65.5	727	9	US-10-097-597-1
7	1337	65.5	727	9	US-10-097-580-1
8	1337	65.5	727	10	US-09-445-023A-1
9	1337	65.5	950	10	US-09-321-987B-4
10	1337	65.5	968	9	US-10-163-316-7
11	1334	65.3	967	12	US-10-105-929-2
12	1181.5	57.9	837	9	US-10-174-590-352
13	1181.5	57.9	837	9	US-10-176-758-352
14	1181.5	57.9	837	9	US-10-173-737-352
15	1181.5	57.9	837	9	US-10-173-706-352
16	1181.5	57.9	837	9	US-10-175-738-352
17	1181.5	57.9	837	9	US-10-175-752-352
18	1181.5	57.9	837	9	US-10-176-482-352
19	1181.5	57.9	837	9	US-10-176-757-352

20	1181.5	57.9	837	9	US-10-176-913-352	Sequence 352, App
21	1181.5	57.9	837	9	US-10-180-552-352	Sequence 352, App
22	1181.5	57.9	837	9	US-10-180-557-352	Sequence 352, App
23	1181.5	57.9	837	9	US-10-173-700-352	Sequence 352, App
24	1181.5	57.9	837	9	US-10-174-572-352	Sequence 352, App
25	1181.5	57.9	837	9	US-10-174-579-352	Sequence 352, App
26	1181.5	57.9	837	9	US-10-174-582-352	Sequence 352, App
27	1181.5	57.9	837	9	US-10-174-588-352	Sequence 352, App
28	1181.5	57.9	837	9	US-10-175-739-352	Sequence 352, App
29	1181.5	57.9	837	9	US-10-175-740-352	Sequence 352, App
30	1181.5	57.9	837	9	US-10-175-743-352	Sequence 352, App
31	1181.5	57.9	837	9	US-10-176-488-352	Sequence 352, App
32	1181.5	57.9	837	9	US-10-176-492-352	Sequence 352, App
33	1181.5	57.9	837	9	US-10-176-747-352	Sequence 352, App
34	1181.5	57.9	837	9	US-10-176-750-352	Sequence 352, App
35	1181.5	57.9	837	9	US-10-176-985-352	Sequence 352, App
36	1181.5	57.9	837	9	US-10-176-987-352	Sequence 352, App
37	1181.5	57.9	837	9	US-10-176-991-352	Sequence 352, App
38	1181.5	57.9	837	9	US-10-176-992-352	Sequence 352, App
39	1181.5	57.9	837	9	US-10-176-993-352	Sequence 352, App
40	1181.5	57.9	837	9	US-10-184-658-352	Sequence 352, App
41	1181.5	57.9	837	9	US-10-173-695-352	Sequence 352, App
42	1181.5	57.9	837	9	US-10-173-697-352	Sequence 352, App
43	1181.5	57.9	837	9	US-10-173-705-352	Sequence 352, App
44	1181.5	57.9	837	9	US-10-174-576-352	Sequence 352, App
45	1181.5	57.9	837	9	US-10-174-585-352	Sequence 352, App

ALIGNMENTS

RESULT 1
US-09-965-631-4
; Sequence 4, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-4

Query Match	100.0%	Score 2042;	DB 10;	Length 950;
Best Local Similarity	100.0%;	Pred. No. 1.8e-169;		
Matches 371;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	FVSIPRYVETLVVADESWKFHGADLEHYLTLTATAARLYRHPHSILNPNIVVVKVLLL	60	
Db	213	FVSIPRYVETLVVADESWKFHGADLEHYLTLTATAARLYRHPHSILNPNIVVVKVLLL	272	
QY	61	RDRDSGPKVTGNAALTLRNFCQWKKLNKVSQKHPYWDTAILETRDLCGATTCDTLGM	120	
Db	273	RDRDSGPKVTGNAALTLRNFCQWKKLNKVSQKHPYWDTAILETRDLCGATTCDTLGM	332	
QY	121	ADVGMCDPKRCSVIEDGLPSAFTTAHELGHVFNPHDNNVKNVCEEVFGKLRANHMSP	180	
Db	333	ADVGMCDPKRCSVIEDGLPSAFTTAHELGHVFNPHDNNVKNVCEEVFGKLRANHMSP	392	
QY	181	TLIQIDRANPWASCSAAITDPLDSGHDCLLDQSPKPSILPDLPGASYTLSSQCELA	240	
Db	393	TLIQIDRANPWASCSAAITDPLDSGHDCLLDQSPKPSILPDLPGASYTLSSQCELA	452	
QY	241	GVGSKPCPYMOTYKTLWCTGKAKGMVCQTRHFWADGTSCGEGKCLCKGACVERHNLK	300	

```
Db 453 GVSGKPCPYMQYCTKLWCTGKAKGQVCTQTHFFPWADGTCGEGKCLKLGACVERHNLK 512
QY 301 HRVDGSKAKWDPYPCSRCTCGGVLARRQCTNPTPANGKYCEGVYKYSRCSNLEPCPS 360
Db 513 HRVDGSKAKWDPYPCSRCTCGGVLARRQCTNPTPANGKYCEGVYKYSRCSNLEPCPS 572
QY 361 SASGKSFREQ 371
Db 573 SASGKSFREQ 583

RESULT 2
US-10-163-316-2
; Sequence 2, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: MP101-025P1RNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2

Query Match 98.9%; Score 2020.5; DB 9; Length 823;
Best Local Similarity 94.2%; Pred. No. 1.1e-167;
Matches 371; Conservative 0; Mismatches 0; Indels 23; Gaps 1;

QY 1 FVSIPIRYVETLVVADESVMVKFHGADLEHYLLTLATARLYRHPSILNPINIVVKVLL 60
Db 213 FVSIPIRYVETLVVADESVMVKFHGADLEHYLLTLATARLYRHPSILNPINIVVKVLL 272
QY 61 RDRDSGPKVTGNAALTNRNFCAMQKLNKVDKHPYWDTAITLFTRODLGATTCDTLGM 120
Db 273 RDRDSGPKVTGNAALTNRNFCAMQKLNKVDKHPYWDTAITLFTRODLGATTCDTLGM 332
QY 121 ADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180
Db 333 ADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 392
QY 181 TLTIQIDRANPWSACSAAIITDFLDGSGDCLLDQPSKPISLPDLPGASYTLSSQCELA 240
Db 393 TLTIQIDRANPWSACSAAIITDFLDGSGDCLLDQPSKPISLPDLPGASYTLSSQCELA 452
QY 241 GVSGKPCPYMQYCTKLWCTGKAKGQVCTQTHFFPWADGTCGEGKCLKLGACVERHNLK 300
Db 453 GVSGKPCPYMQYCTKLWCTGKAKGQVCTQTHFFPWADGTCGEGKCLKLGACVERHNLK 512
QY 301 HR-----VDSWAKWDPYPCSRCTCGGVLARRQCTNPTPA 337
Db 513 HRPTDIISPQKLLRLPNGLHTTQVDSWAKWDPYPCSRCTCGGVLARRQCTNPTPA 572
QY 338 NGKYCEGVYKYSRCSNLEPCPSASGKSFREQ 371
Db 573 NGKYCEGVYKYSRCSNLEPCPSASGKSFREQ 606
```

```
RESULT 3
US-10-097-597-12
; Sequence 12, Application US/10097597
; Publication No. US2003002352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiiji
```

```
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-097-597-12

Query Match 65.6%; Score 1339; DB 9; Length 727;
Best Local Similarity 63.2%; Pred. No. 2e-108;
Matches 237; Conservative 52; Mismatches 80; Indels 6; Gaps 4;

QY 1 FVSIPIRYVETLVVADESVMVKFHGADLEHYLLTLATARLYRHPSILNPINIVVKVLL 60
Db 13 FVSIPIRYVETLVVADESVMVKFHGADLEHYLLTLATARLYRHPSILNPINIVVKVLL 72
QY 61 RDRDSGPKVTGNAALTNRNFCAMQKLNKVDKHPYWDTAITLFTRODLGATTCDTLGM 120
Db 73 YEEQKPEVTSNAALTNRNFCAMQKLNKVDKHPYWDTAITLFTRODLGATTCDTLGM 132
QY 121 ADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180
Db 133 ADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 192
QY 181 TLTIQIDRANPWSACSAAIITDFLDGSGDCLLDQPSKPISLPDLPGASYTLSSQCELA 240
Db 193 MUSSLDHSPQSPSCSAYMVTSLFNGHGECLMDKPNKIPKLPDLPGTLYDANRQCQFT 252
QY 241 GVSGKPCPYMQYCTKLWCTGKAKGQVCTQTHFFPWADGTCGEGKCLKLGACVERHNL 299
Db 253 GEESKHCDDAATCTTACTGCTSGLLVCTQTHFFPWADGTCGEGKWCVCVKNKTD 311
QY 300 KH---RVDSWAKWDPYPCSRCTCGGVLARRQCTNPTPANGKYCEGVYKYSRCSNLE 356
Db 312 KHFAFVHSGWPGWPGWDCSRTCGGVQVYTMRECDNVPKNGKGYCEGKRVYRSCNIE 371
QY 357 PCPSSASGKSFREQ 371
Db 372 DCPDN-NKGTFFREQ 385

RESULT 4
US-10-097-580-12
; Sequence 12, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiiji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
```

; PRIOR APPLICATION NUMBER: 09/445,023
 ; PRIOR FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 727
 ; TYPE: PRT
 ; ORGANISM: Mus sp.
 US-10-097-580-12

Query Match 65.6%; Score 1339; DB 9; Length 727;
 Best Local Similarity 63.2%; Pred. No. 2e-108;
 Matches 237; Conservative 52; Mismatches 80; Indels 6; Gaps 4;
 QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLLAARLYRHPHSILNPINIVVVKVLL 60
 DB 13 FVSSPRIVETMLVADQSMADFHGSLKHYLLTFLSVAARFYKHPHSIRNSISLVVVKILVI 72
 QY 61 RDRSGPKVTGNAALTNRNFCAMOKKLNKVS DKHPYWDTAILFTRODLGATTCDTLGM 120
 DB 73 YEQKGPEVTSNAALTNRNFCAMOKKLNKVS DKHPYWDTAILFTRODLGATTCDTLGM 132
 QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNKVCYEEVFGKLRANHMSP 180
 DB 133 ADVGTVCDPSPSCSVIEDDGLQAFTTAHELGHVFNPHDDAKHACASLVGSDSHLMA 192
 QY 181 TLIQIDRANPWSACSAIITDFLDGSGDCLLDQSPKPSISLPEDLPASVYTLSSQCEIAF 240
 DB 193 MSLSDHSPWSPSCSAIYMTSFLDNHGHCECLMDKPQNIKLPDLPGLTYDANRQCQFTF 252
 QY 241 GVSGKPCP-YMQYCTKLWCTGKAKGOMVCTRHFPWADGTSCGEGKLCGLKACVERHN 299
 DB 253 GEESKHCPPAASTCTTLCWCTSGGLLVCTQKHPFPWADGTSCGEGKVCVKNKYTDM- 311
 QY 300 KH---RVDSGWAKWDYPCSRCTCGGVQLARRQCTNPTPANGKYCEGVRYKRSNLE 356
 DB 312 KHFPATPVHSGWPGWPGWDCSRTCGGVQYTMRECDNPVPKNGGKYCEGKRVYRSCNIE 371
 QY 357 PCPSSASGKSFREQ 371
 DB 372 DCPDN-NGKTFREQ 385

RESULT 5
 US-09-445-023A-12
 ; Sequence 12, Application US/09445023A
 ; Patent No. US20020119167A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirose, Kunitaka
 ; APPLICANT: Inoguchi, Eiichi
 ; APPLICANT: Hakozaaki, Michinori
 ; APPLICANT: Ishioke, Keiko
 ; APPLICANT: Ishida, Yukako
 ; APPLICANT: Matsushima, Kouji
 ; APPLICANT: Kuno, Kouji
 ; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
 ; FILE REFERENCE: Q57092
 ; CURRENT APPLICATION NUMBER: US/09/445,023A
 ; PRIOR FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 727
 ; TYPE: PRT
 ; ORGANISM: Mus sp.
 US-09-445-023A-12

Query Match 65.6%; Score 1339; DB 10; Length 727;

Best Local Similarity 63.2%; Pred. No. 2e-108;
 Matches 237; Conservative 52; Mismatches 80; Indels 6; Gaps 4;
 QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLLAARLYRHPHSILNPINIVVVKVLL 60
 DB 13 FVSSPRIVETMLVADQSMADFHGSLKHYLLTFLSVAARFYKHPHSIRNSISLVVVKILVI 72
 QY 61 RDRSGPKVTGNAALTNRNFCAMOKKLNKVS DKHPYWDTAILFTRODLGATTCDTLGM 120
 DB 73 YEQKGPEVTSNAALTNRNFCAMOKKLNKVS DKHPYWDTAILFTRODLGATTCDTLGM 132
 QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNKVCYEEVFGKLRANHMSP 180
 DB 133 ADVGTVCDPSPSCSVIEDDGLQAFTTAHELGHVFNPHDDAKHACASLVGSDSHLMA 192
 QY 181 TLIQIDRANPWSACSAIITDFLDGSGDCLLDQSPKPSISLPEDLPASVYTLSSQCEIAF 240
 DB 193 MSLSDHSPWSPSCSAIYMTSFLDNHGHCECLMDKPQNIKLPDLPGLTYDANRQCQFTF 252
 QY 241 GVSGKPCP-YMQYCTKLWCTGKAKGOMVCTRHFPWADGTSCGEGKLCGLKACVERHN 299
 DB 253 GEESKHCPPAASTCTTLCWCTSGGLLVCTQKHPFPWADGTSCGEGKVCVKNKYTDM- 311
 QY 300 KH---RVDSGWAKWDYPCSRCTCGGVQLARRQCTNPTPANGKYCEGVRYKRSNLE 356
 DB 312 KHFPATPVHSGWPGWPGWDCSRTCGGVQYTMRECDNPVPKNGGKYCEGKRVYRSCNIE 371
 QY 357 PCPSSASGKSFREQ 371
 DB 372 DCPDN-NGKTFREQ 385

RESULT 6
 US-10-097-597-1
 ; Sequence 1, Application US/10097597
 ; Publication No. US20030022352A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirose, Kunitaka
 ; APPLICANT: Inoguchi, Eiichi
 ; APPLICANT: Hakozaaki, Michinori
 ; APPLICANT: Ishioke, Keiko
 ; APPLICANT: Ishida, Yukako
 ; APPLICANT: Matsushima, Kouji
 ; APPLICANT: Kuno, Kouji
 ; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
 ; TITLE OF INVENTION: pharmaceutical
 ; FILE REFERENCE: Q57092
 ; CURRENT APPLICATION NUMBER: US/10/097,597
 ; CURRENT FILING DATE: 2002-03-15
 ; PRIOR APPLICATION NUMBER: 09/445,023
 ; PRIOR FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 727
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-097-597-1

Query Match 65.5%; Score 1337; DB 9; Length 727;
 Best Local Similarity 63.5%; Pred. No. 3e-108;
 Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;
 QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLLAARLYRHPHSILNPINIVVVKVLL 60
 DB 13 FVSSPRIVETMLVADQSMADFHGSLKHYLLTFLSVAARLYRHPHSIRNSISLVVVKILVI 72
 QY 61 RDRSGPKVTGNAALTNRNFCAMOKKLNKVS DKHPYWDTAILFTRODLGATTCDTLGM 120
 DB 73 HDEQKGPVTSNAALTNRNFCAMOKKLNKVS DKHPYWDTAILFTRODLGATTCDTLGM 132


```

; Sequence 4, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Kimble, Judith E
; APPLICANT: Blleloch, Robert H
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296 95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/129,023
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Murine
US-09-321-987B-4

Query Match      65.5%; Score 1337; DB 10; Length 950;
Best Local Similarity 63.2%; Pred. No. 4.2e-108;
Matches 237; Conservative 51; Mismatches 81; Indels 6; Gaps 4;

Qy 1 FVSPRYVETLWVADESVMVKFHGADLEHYLLTLLATAAARLYRHPHSILNPINIVVVKVLL 60
   ||| |||||:::||||| |||: ||:|||||: ||| ||:|||| |:::|||||:::
Db 237 FVSPRYVETMLVADQSMADFHSGLKHLYLLTFSVAARFYKHPHSIRNSISLVVVKILVI 296
   ||| |||||:::||||| |||: ||:|||||: ||| ||:|||| |:::|||||:::

Qy 61 RDRDSGPKVTGNAALTLRNFCWAKKLNKVSDDKHPYWDTAILEFTRODLGGATTCDTLGM 120
   ||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 YEEQKGPEVTSNAALTLRNFCWQKQNSPSDRDPEHYDTAILFTRODLGSHCTDLGM 356
   ||| |||||:::||||| |||: ||:|||||: ||| ||:|||| |:::|||||:::

Qy 121 ADVCTMCDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDNKVCVEEYFGKLRANHMSP 180
   ||| |||||:::||||| |||: ||:|||||: ||| ||:|||| |:::|||||:::
Db 357 ADVGTCDPKRSCSVIEDGLQAFTTAHELGHVFNPHDDAKHCASLNGVTGDSHLMAS 416
   ||| |||||:::||||| |||: ||:|||||: ||| ||:|||| |:::|||||:::

Qy 181 TLIQIDRANPWSGASAIITFDLSGHGDCDLLDQSPKIPSLPEDLPGASYTLSQOCELAF 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 MLSLSDHSQWPSPCSAYMVTSLDNGHGECMLMDRFPQPIKLPDLPTGLYDANRQCQFTF 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 241 GVSGKPCP-QYQYTKLWCTGKAKQWVCOTRHPFPWADGTCGREGKLCGLKACGYVERHNL 299
   || || || || || || || || || || || || || || || || || || || || ||
Db 477 GEESKHCPDAASTCTTLWCTGTSGLLVLCOTKHFPWADGTCGREGKVCVKNKTDMM- 535
   || || || || || || || || || || || || || || || || || || || || ||

Qy 300 KH---RYDGGSWAKWDPYPCSRTCGGGVQLARRCOTNPTPANGGKYCEGVRYKRSCLNLE 356
   || || || || || || || || || || || || || || || || || || || || ||
Db 536 KHFAFPYHGSWGPWPGWDCSRTCGGGVQVYTMRECDNPVPANGGKYCEGARVRYRSCNIE 595
   || || || || || || || || || || || || || || || || || || || || ||

Qy 357 PCPSSASGKSFREQ 371
   || : ||: |||||
Db 596 DCPDN-NGKTFREQ 609
   || : ||: |||||

RESULT 10
US-10-163-316-7
; Sequence 7, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: Therefor
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus

```

```
Db 373 ADGVTCDPSRSCSVIEDDGLQAAFTTAHELGHVFNMFPHDDAKOCASLNGVNDSHMMAS 432
QY 181 TLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKPSISLPEDLPASGYTLSSQOCELAF 240
Db 433 MSLNLDHSPWSPSCSAYMITSLDNGHGECLEMDKPNPQLPGLPGTSTYDANRQCQTFE 492
QY 241 GVGSKPCP-YMOYCTKLMCTGKAKOMVQOTRHPFPWADGTSCGEGKLCCLKGACVERHNLN 299
Db 493 GDSKHCPCDAASTCTLMCTGSGGVLVCOTKHFPPWADGTSCGEGKWCINKCKVNTD-R 551
QY 300 KH---RVDGSWAKWPYGPCSRTCGGVLQARRQCTNPTPANGGKYCEGVVRYKRSNCLE 356
Db 552 KHFDTPFHSGWCMGWPWGDCSRTCGGVQYTMRECDNPVPKNGKCYCEGKRVRYRSCNLE 611
QY 357 PCPSSASGKSFRREQ 371
Db 612 DCPDN-NGKTFREQ 625

RESULT 12
US-10-174-590-352
; Sequence 352, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-352

Query Match 57.9%; Score 1181.5; DB 9; Length 837;
Best Local Similarity 55.5%; Pred. No. 1.2e-94;
Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSIPRYVETLVVADESVMKPHGADLEHYLLTLATAARLYRHPHSILNPINIVVVKVLL 60
Db 213 FASLSRFVETLVVADDKMAAFHAGLKKRYLLTVMAAAKAFKHSIRNPVSLVTVRLVIL 272
QY 61 RDRDSGPKVTGNAALTLRNFCWQKLNKVS DKHPEYWDTAITLFTRODLGATTCDTLGM 120
Db 273 GSGEGPQVGPAAQTLRSFCAWQRLNTPEDSGPDHFDTAITLFTRODLGCVSTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMFPHDDAKOCASLNGVNDSHMMAS 179
Db 333 ADVGTVCDFPARSCAIVEDDGLQSAFTAHELGHVFNMLHDNSKPCISLNGPLSTRHVMA 392
QY 180 PTLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKPSISLPEDLPASGYTLSSQOCELA 239
Db 393 PVNAHVDPPEPSPCSARFIDFLDNGYGHCLLDKPEAPLHPVTFPGKYDADROCOLT 452
QY 240 FGVGSKPCPYM-QYCTKLMCTGKAKOMVQOTRHPFPWADGTSCGEGKLCCLKGACVERHN 298
Db 453 FGPDSRHCPQLPPPCAAALWCSGHLNGLHMACQTKHSPWADGTPCGPAQACMGRCRLHMDQL 512
QY 299 NKHRVD--GSAWKWPYGPCSRTCGGVLQARRQCTNPTPANGGKYCEGVVRYKRSNCLE 356
Db 513 QDFNTPQAGGWPWGDCSRTCGGVQYFSSRDCTRPVPNGGKYCEGRTRFRSCNTE 572

RESULT 14
US-10-175-737-352
```

```
QY 299 NKHRVD--GSAWKWPYGPCSRTCGGVLQARRQCTNPTPANGGKYCEGVVRYKRSNCLE 356
Db 513 QDFNTPQAGGWPWGDCSRTCGGVQYFSSRDCTRPVPNGGKYCEGRTRFRSCNTE 572
QY 357 PCPSSASGKSFRREQ 371
Db 573 DCP-TGSALTFRREQ 586

RESULT 13
US-10-176-758-352
; Sequence 352, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-352

Query Match 57.9%; Score 1181.5; DB 9; Length 837;
Best Local Similarity 55.5%; Pred. No. 1.2e-94;
Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSIPRYVETLVVADESVMKPHGADLEHYLLTLATAARLYRHPHSILNPINIVVVKVLL 60
Db 213 FASLSRFVETLVVADDKMAAFHAGLKKRYLLTVMAAAKAFKHSIRNPVSLVTVRLVIL 272
QY 61 RDRDSGPKVTGNAALTLRNFCWQKLNKVS DKHPEYWDTAITLFTRODLGATTCDTLGM 120
Db 273 GSGEGPQVGPAAQTLRSFCAWQRLNTPEDSGPDHFDTAITLFTRODLGCVSTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMFPHDDAKOCASLNGVNDSHMMAS 179
Db 333 ADVGTVCDFPARSCAIVEDDGLQSAFTAHELGHVFNMLHDNSKPCISLNGPLSTRHVMA 392
QY 180 PTLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKPSISLPEDLPASGYTLSSQOCELA 239
Db 393 PVNAHVDPPEPSPCSARFIDFLDNGYGHCLLDKPEAPLHPVTFPGKYDADROCOLT 452
QY 240 FGVGSKPCPYM-QYCTKLMCTGKAKOMVQOTRHPFPWADGTSCGEGKLCCLKGACVERHN 298
Db 453 FGPDSRHCPQLPPPCAAALWCSGHLNGLHMACQTKHSPWADGTPCGPAQACMGRCRLHMDQL 512
QY 299 NKHRVD--GSAWKWPYGPCSRTCGGVLQARRQCTNPTPANGGKYCEGVVRYKRSNCLE 356
Db 513 QDFNTPQAGGWPWGDCSRTCGGVQYFSSRDCTRPVPNGGKYCEGRTRFRSCNTE 572

RESULT 14
US-10-175-737-352
```

```
; Sequence 352, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10175.737
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-352

Query Match 57.9%; Score 1181.5; DB 9; Length 837;
Best Local Similarity 55.5%; Pred. No. 1.2e-94;
Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPHSILNINIVVVKVLL 60
Db 213 FASLSRFVETLVVADDKAAAFHAGLKYLLTVMAAAKAFKHPHSIRNPVSLVTVRLVIL 272
QY 61 RDRDSGPKVTGNAALTNRNFCANQKLNKYSKHPYWDTAIFTRQDLCGATTCDTLGM 120
Db 273 GSGEEGPQVGSAAQTLRSFCAWQRLNTPEDSGPDHFDTAIFTRQDLCGVSTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFKL-RANHMS 179
Db 333 ADVGTVCDPARSCAIVEDDGLQSAFTAHELGHVFNHMLHDSKPCISLNGPLSTRHVA 392
QY 180 PTLIQIDRANPWSACSAAITDFDSHGDCILLDQSPKPSILPEDLPASYYTLSSQOCELA 239
Db 393 PVMAHVDPEEPWSPCSARFITDFLDNGYGHCLLDKPEAPLHPVTFPGKDYDADRQCLT 452
QY 240 FGVGSKPCPYM-OYCTKLWCTGKAKGQWVCOTRHPFWADGTSCGEGKLCCLKGACVERHNL 298
Db 453 FGPDNRHCPQLPPPCAAALWCSGHLNGHAMCOTKHSPWADGTCGPAQACMGGRCLHMDQL 512
QY 299 NKHRVD--GSNAKWDPYPCSRCTCGGVQLARRQCTNPTPANGKYCEGVVRYKSCNLE 356
Db 513 QDFNIPQAGGWPWGPWGDSCRTCGGVQFSSRDCTRPVPRNGGKYCEGRRTRFRSCNTE 572
QY 357 PCPSSASGKSFREQ 371
Db 573 DCP-TGSALTTFREQ 586

RESULT 15
US-10-173-706-352
; Sequence 352, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
```

```
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10173.706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-352

Query Match 57.9%; Score 1181.5; DB 9; Length 837;
Best Local Similarity 55.5%; Pred. No. 1.2e-94;
Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPHSILNINIVVVKVLL 60
Db 213 FASLSRFVETLVVADDKAAAFHAGLKYLLTVMAAAKAFKHPHSIRNPVSLVTVRLVIL 272
QY 61 RDRDSGPKVTGNAALTNRNFCANQKLNKYSKHPYWDTAIFTRQDLCGATTCDTLGM 120
Db 273 GSGEEGPQVGSAAQTLRSFCAWQRLNTPEDSGPDHFDTAIFTRQDLCGVSTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFKL-RANHMS 179
Db 333 ADVGTVCDPARSCAIVEDDGLQSAFTAHELGHVFNHMLHDSKPCISLNGPLSTRHVA 392
QY 180 PTLIQIDRANPWSACSAAITDFDSHGDCILLDQSPKPSILPEDLPASYYTLSSQOCELA 239
Db 393 PVMAHVDPEEPWSPCSARFITDFLDNGYGHCLLDKPEAPLHPVTFPGKDYDADRQCLT 452
QY 240 FGVGSKPCPYM-OYCTKLWCTGKAKGQWVCOTRHPFWADGTSCGEGKLCCLKGACVERHNL 298
Db 453 FGPDNRHCPQLPPPCAAALWCSGHLNGHAMCOTKHSPWADGTCGPAQACMGGRCLHMDQL 512
QY 299 NKHRVD--GSNAKWDPYPCSRCTCGGVQLARRQCTNPTPANGKYCEGVVRYKSCNLE 356
Db 513 QDFNIPQAGGWPWGPWGDSCRTCGGVQFSSRDCTRPVPRNGGKYCEGRRTRFRSCNTE 572
QY 357 PCPSSASGKSFREQ 371
Db 573 DCP-TGSALTTFREQ 586
```

Search completed: April 29, 2003, 17:21:49
Job time : 12.2469 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 8.58125 Seconds
(without alignments)
4156.253 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_583

Perfect score: 2042

Sequence: 1 FVSIPIRYVETLVVADESVMVK.....SCNLEPCPSASGSKSPREQ 371

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1337	65.5	951	2 T00017	gene ADAMTS-1 prot
2	1182.5	57.9	837	2 T00355	hypothetical prote
3	923.5	45.2	2165	2 T21371	hypothetical prote
4	684	33.5	550	2 T47158	hypothetical prote
5	628.5	30.8	1205	2 T18517	procollagen N-endo
6	393	19.2	860	2 T16892	hypothetical prote
7	377.5	18.5	1444	2 T18856	angiogenesis inhib
8	297.5	14.6	957	2 T15976	hypothetical prote
9	233.5	11.4	903	2 S60257	meltrin alpha - mo
10	231.5	11.3	508	2 T22836	hypothetical prote
11	229.5	11.2	1170	1 TSHUP1	thrombospondin 1 p
12	224.5	11.0	1170	2 A40558	thrombospondin 1 p
13	220	10.8	571	2 S24789	thrombospondin 1 p
14	218.5	10.7	419	2 S41607	thrombospondin 1 p
15	217.5	10.7	419	2 A59414	thrombospondin 1 p
16	213	10.4	549	2 A8169	thrombospondin 1 p
17	210	10.3	416	2 A37877	thrombospondin 1 p
18	209.5	10.3	617	2 S48160	thrombospondin 1 p
19	208.5	10.2	609	2 S55270	thrombospondin 1 p
20	201.5	9.9	407	2 S66260	thrombospondin 1 p
21	199	9.7	826	2 A60385	thrombospondin 1 p
22	197.5	9.7	411	1 HYSNFA	thrombospondin 1 p
23	197	9.6	480	1 A30065	thrombospondin 1 p
24	196.5	9.6	814	2 G03290	thrombospondin 1 p
25	195.5	9.6	203	2 A59421	thrombospondin 1 p
26	194	9.5	1074	2 JCS928	thrombospondin 1 p
27	193.5	9.5	481	2 JCS432	thrombospondin 1 p
28	191.5	9.4	202	2 JC2550	thrombospondin 1 p
29	191	9.4	478	2 JC4880	thrombospondin 1 p

30	189	9.3	429	2 A42972	coagulation factor
31	187.5	9.2	414	2 S41609	coagulation factor
32	184.5	9.0	481	2 S43125	coagulation factor
33	181.5	8.9	414	1 HYSAC	coagulation factor
34	181	8.9	600	2 I49281	coagulation factor
35	180.5	8.8	952	2 T18900	coagulation factor
36	180	8.8	1172	2 A42587	coagulation factor
37	179.5	8.8	200	2 S15111	coagulation factor
38	177	8.7	1572	2 T00027	coagulation factor
39	174.5	8.5	1178	1 A39804	coagulation factor
40	173	8.5	478	2 A43296	coagulation factor
41	170.5	8.3	469	1 S29126	coagulation factor
42	170.5	8.3	1172	1 TSHUP2	coagulation factor
43	170	8.3	202	1 HYSR	coagulation factor
44	170	8.3	437	2 S05478	coagulation factor
45	169.5	8.3	1584	2 T00026	coagulation factor

ALIGNMENTS

RESULT 1

T00017

gene ADAMTS-1 protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00017

R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.

Genomics 46, 466-471, 1997

A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1

A:Reference number: Z14055; MUID:98110583; PMID:9441751

A:Accession: T00017

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-951 <KUN>

A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057

A:Experimental source: strain 129SVJ

C:Genetics:

A:Gene: ADAMTS-1

A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2

A:Superfamily: thrombospondin type 1 repeat homology

F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 65.5%; Score 1337; DB 2; Length 951;

Best Local Similarity 63.2%; Pred No. 3.9e-97;

Matches 237; Conservative 51; Mismatches 81; Indels 6; Gaps 4;

QY 1 FVSIPIRYVETLVVADESVMVKFHGADLEHYLLTLLATAARLYRHPISILAPINIVVVKVLL 60

Db 237 FVSSPIRYVETLVVADESVMVKFHGADLEHYLLTLLATAARLYRHPISILAPINIVVVKVLLVI 296

QY 61 RDRSGPKVTGNAALTLRNFCAQWKLKVKDHPYWDTAILFTQDLCGATTCDTLGM 120

Db 297 YEEQKGEVTSNAALTLRNFCAQWKLKVKDHPYWDTAILFTQDLCGATTCDTLGM 356

QY 121 ADVGTMDPKRCSVIEDDGLPSAFTTAELGHVFMHPDHNKVCVEEYFGKLRANHMSP 180

Db 357 ADVGTMDPKRCSVIEDDGLPSAFTTAELGHVFMHPDHNKVCVEEYFGKLRANHMSP 416

QY 181 TLIIQIDRANPWSACSAIITDFDSHGDCDLDQSPKISLPEDLPASYYTISQCELA 240

Db 417 MLSSLDHSPWSPCSAYMYVTSFLDNGHGBCLMDKPNPKLPDLPGTLYDANRQCQTF 476

QY 241 GYSGRCP- YMOYCYKFLACTGAKQWQVQTRHFWADGTSCGEGKCLKGCACVERHNL 299

Db 477 GEESKHCPCDAATCTTLLTLLTLLTLLTLLTLLTLLTLLTLLTLLTLLTLLTLLTLL 535

QY 300 KH- --RVDSWAKWDPYGPSCSRTCGGVQLARRQCTNPTPANGKVCYGVGVKYSNCLE 356

Db 536 KHFAFPVHGSWGPWGPWGDSCSTCGGVQYTHRECDNPVPKNGKVCYGVGVKYSNCLE 595

QY 357 PCPSSASGSKSPREQ 371

|| : ||:|||||

Db 596 DCPDN-NGKTFREQ 609

RESULT 2

T00355

hypothetical protein KIAA0688 - human

C:Species: Homo sapiens (man)

C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000

C:Accession: T00355

R:Shikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.

DNA Res. 5, 169-176, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete

A:Reference number: Z14142; MUID:98403880; PMID:9734811

A:Accession: T00355

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-837 <ISH>

A:CROSS-references: EMBL:AB014588; NID:g3327189; PIDN:BAA31663.1; PID:g3327190

A:Experimental source: brain

C:Genetics:

A:Gene: KIAA0688

C:Superfamily: thrombospondin type 1 repeat homology

F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 57.9%; Score 1182.5; DB 2; Length 837;

Best Local Similarity 55.5%; Pred. No. 4.7e-85;

Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSTPRVETLVVADESMVKFHGADLEHYLLTLTAARLYRHPISILNINIVVKKVLL 60

Db 213 FASLSREVELVWADDKAAAFHAGLAKRYLLTVMAAAKAFKPSIRNPVSLVTVLVL 272

QY 61 RDRSGPKVTGNAALTNRNFCAMOKKLVKSDKHPEYWDTAITFTRODLGATTCDTLGM 120

Db 273 GSGEGGQVGPFAQAOTURSCAMGUNTPEDDPDHFDTAITFTRODLGCGVCTDLGM 332

QY 121 ADVGMDCKRRSCSVIEDDGLPSAFTTAHELGHVFNPHDVKVCEVFGKL-RANRMS 179

Db 333 ADVGTVCDPARSCAIVEDDGLQSAFTAAHELGHVFNHLDNSKPCISLNGPLSTSRVMA 392

QY 180 PTLTQIDRANPWSACSAIITDFLDLSDGHCCLLDQPSKPSLPEDLPGASYTLSSQCELA 239

Db 393 PVMAHVDPPEPWPSCARFITDFLDNGYGHCLLDKPEAPLHPVTPFGKYDADROCOLT 452

QY 240 FGVGSKPCPYM-QYCTKLWCTGKAKGMVQTRHFPWADTSGEGKCLKLGACVERHNL 298

Db 453 FGPDNRHCPQLPPCAALWCSSLHNGHMCOTKHSFWADGTPCGPAQACNGRCLHMDQL 512

QY 299 NKHRVD--GSAKWDPYGPSCRTCCGGVQLARQCTNPTANGKYCEGVRYVYRSCNLE 356

Db 513 QDFNIPQAGGPGWPGWDCSRTCCGGVQFSSRDCTRPVPRNGKYCEGRRTFRSCNTE 572

QY 357 PCPSASGSKSFREQ 371

Db 573 DCP-TGSALTFRREQ 586

RESULT 3

T21371

hypothetical protein F25H8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T21371; T24896

R:Gajadaty, S.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19413

A:Accession: T21371

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2165 <NID>

A:CROSS-references: EMBL:Z693360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3

A:Experimental source: clone F25H8

R:Gajadaty, S.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19949

A:Accession: T24896

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2165 <NID>

A:CROSS-references: EMBL:Z693361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3

A:Experimental source: clone T13H10

C:Genetics:

A:Gene: CESP:F25H8.3

A:Map position: 4

A:introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1;

Query Match 45.2%; Score 923.5; DB 2; Length 2165;

Best Local Similarity 45.4%; Pred. No. 2.9e-64;

Matches 177; Conservative 60; Mismatches 126; Indels 27; Gaps 7;

QY 7 YVETLVVADESMVKFHGADLEHYLLTLTAARLYRHPISILNINIVVKKVLLLRDRSG 66

Db 281 YVEVLVWADTKMYEYHGRSLEDYVLTFTVASIYRHQSLRASINIVVVKLVLTENAG 340

QY 67 PKVTGNAALTNRNFCAMOKKLVKSDKHPEYWDTAITFTRODLGAT-TCDTLGMADVGT 125

Db 341 PRITONAQTLDQFCRWQOYNDPDDSSVQHDVAILLTRKDICRSQCKGKDTLGLAELGT 400

QY 126 MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDVKVCEV-----FGKL 172

Db 401 MCDMQKCAIIEDNGLSAAFTTAHELGHVFSIPHDERKCYMPVNVKVKCFQSTKFDKT 460

QY 173 RAN---HMSPPLTQIDRANPWSACSAIITDFLDLSDHG--DCLLDQPSKPSISUPE---- 223

Db 461 QPNQNFHIMAPTLEYNTHPWSWSPCSAGMLERLENNGRQTCFLDQPVVRYEDVFR 520

QY 224 DLPGASYTLSSQCELAFAFGVSGKPCPYMCTKLWCTGKAKGMVQTRHFPWADTSGCE 283

Db 521 DDFGKKYDAHQCKVFGVPASELCYMPCTCRWLWCATYFGSQMGCRTHMPWADGTPCDE 580

QY 284 GK--LCLKGACVERINLKHVRVDGSAKWDPYGPSCRTCCGGVQLARQCTNPTANGGK 341

Db 581 SRSMFCHGACVRLAPESLTKIDGQWDRWSWGECSRTCCGGVQKGLRDCDSPKRNNGK 640

QY 342 YCEGVRYVYRSCNLEPCPSASGSKSFREQ 371

Db 641 YCVQGERYRSCNTQECPWDT--QYREVQ 668

RESULT 4

T47158

hypothetical protein DKFp762C110.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47158

R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24379

A:Accession: T47158

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-550 <AAA>

A:CROSS-references: EMBL:AL162080

A:Experimental source: adult melanoma (Mewo cell line); clone DKFp762C110

C:Genetics:

A:Note: DKFp762C110.1

Query Match 33.5%; Score 684; DB 2; Length 550;

Best Local Similarity 59.2%; Pred. No. 4.4e-46;

Matches 119; Conservative 25; Mismatches 51; Indels 6; Gaps 4;

QY 175 NHMSPPLTQIDRANPWSACSAIITDFLDLSDGHCCLLDQPSKPSLPEDLPGASYTLSSQ 234

Db 10 SHMASMLNLDHDSQWPCSAIYMITSFLDNGHGECLMDKPNQPIQLPGDLPGTSYDANR 69

QY 235 QCELAFAFGVSGKPCPYMCTKLWCTGKAKGMVQTRHFPWADTSGCEGKCLKLGACV 293

Accession: T16892
Bentley, D.
Submitted to the EMBL Data Library, December 1995
Description: The sequence of *C. elegans* cosmid T19D2.
Reference number: Z18599
Accession: T16892
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-860 <BEN>
Cross-references: EMBL:U42846; NID:g1125809; PID:g1125810; PIDN:AAA83600.1; CESP:T16892.1
Genetics:
Gene: CESP:T19D2.1
Exons: 56/3; 96/1; 171/3; 199/3; 245/2; 293/2; 335/3; 395/3; 426/3; 484/2; 505/3; 566/3; 600/3; 630/3; 660/3; 690/3; 720/3; 750/3; 780/3; 810/3; 840/3; 870/3; 900/3; 930/3; 960/3; 990/3; 1020/3; 1050/3; 1080/3; 1110/3; 1140/3; 1170/3; 1200/3; 1230/3; 1260/3; 1290/3; 1320/3; 1350/3; 1380/3; 1410/3; 1440/3; 1470/3; 1500/3; 1530/3; 1560/3; 1590/3; 1620/3; 1650/3; 1680/3; 1710/3; 1740/3; 1770/3; 1800/3; 1830/3; 1860/3; 1890/3; 1920/3; 1950/3; 1980/3; 2010/3; 2040/3; 2070/3; 2100/3; 2130/3; 2160/3; 2190/3; 2220/3; 2250/3; 2280/3; 2310/3; 2340/3; 2370/3; 2400/3; 2430/3; 2460/3; 2490/3; 2520/3; 2550/3; 2580/3; 2610/3; 2640/3; 2670/3; 2700/3; 2730/3; 2760/3; 2790/3; 2820/3; 2850/3; 2880/3; 2910/3; 2940/3; 2970/3; 3000/3; 3030/3; 3060/3; 3090/3; 3120/3; 3150/3; 3180/3; 3210/3; 3240/3; 3270/3; 3300/3; 3330/3; 3360/3; 3390/3; 3420/3; 3450/3; 3480/3; 3510/3; 3540/3; 3570/3; 3600/3; 3630/3; 3660/3; 3690/3; 3720/3; 3750/3; 3780/3; 3810/3; 3840/3; 3870/3; 3900/3; 3930/3; 3960/3; 3990/3; 4020/3; 4050/3; 4080/3; 4110/3; 4140/3; 4170/3; 4200/3; 4230/3; 4260/3; 4290/3; 4320/3; 4350/3; 4380/3; 4410/3; 4440/3; 4470/3; 4500/3; 4530/3; 4560/3; 4590/3; 4620/3; 4650/3; 4680/3; 4710/3; 4740/3; 4770/3; 4800/3; 4830/3; 4860/3; 4890/3; 4920/3; 4950/3; 4980/3; 5010/3; 5040/3; 5070/3; 5100/3; 5130/3; 5160/3; 5190/3; 5220/3; 5250/3; 5280/3; 5310/3; 5340/3; 5370/3; 5400/3; 5430/3; 5460/3; 5490/3; 5520/3; 5550/3; 5580/3; 5610/3; 5640/3; 5670/3; 5700/3; 5730/3; 5760/3; 5790/3; 5820/3; 5850/3; 5880/3; 5910/3; 5940/3; 5970/3; 6000/3; 6030/3; 6060/3; 6090/3; 6120/3; 6150/3; 6180/3; 6210/3; 6240/3; 6270/3; 6300/3; 6330/3; 6360/3; 6390/3; 6420/3; 6450/3; 6480/3; 6510/3; 6540/3; 6570/3; 6600/3; 6630/3; 6660/3; 6690/3; 6720/3; 6750/3; 6780/3; 6810/3; 6840/3; 6870/3; 6900/3; 6930/3; 6960/3; 6990/3; 7020/3; 7050/3; 7080/3; 7110/3; 7140/3; 7170/3; 7200/3; 7230/3; 7260/3; 7290/3; 7320/3; 7350/3; 7380/3; 7410/3; 7440/3; 7470/3; 7500/3; 7530/3; 7560/3; 7590/3; 7620/3; 7650/3; 7680/3; 7710/3; 7740/3; 7770/3; 7800/3; 7830/3; 7860/3; 7890/3; 7920/3; 7950/3; 7980/3; 8010/3; 8040/3; 8070/3; 8100/3; 8130/3; 8160/3; 8190/3; 8220/3; 8250/3; 8280/3; 8310/3; 8340/3; 8370/3; 8400/3; 8430/3; 8460/3; 8490/3; 8520/3; 8550/3; 8580/3; 8610/3; 8640/3; 8670/3; 8700/3; 8730/3; 8760/3; 8790/3; 8820/3; 8850/3; 8880/3; 8910/3; 8940/3; 8970/3; 9000/3; 9030/3; 9060/3; 9090/3; 9120/3; 9150/3; 9180/3; 9210/3; 9240/3; 9270/3; 9300/3; 9330/3; 9360/3; 9390/3; 9420/3; 9450/3; 9480/3; 9510/3; 9540/3; 9570/3; 9600/3; 9630/3; 9660/3; 9690/3; 9720/3; 9750/3; 9780/3; 9810/3; 9840/3; 9870/3; 9900/3; 9930/3; 9960/3; 9990/3; 10020/3; 10050/3; 10080/3; 10110/3; 10140/3; 10170/3; 10200/3; 10230/3; 10260/3; 10290/3; 10320/3; 10350/3; 10380/3; 10410/3; 10440/3; 10470/3; 10500/3; 10530/3; 10560/3; 10590/3; 10620/3; 10650/3; 10680/3; 10710/3; 10740/3; 10770/3; 10800/3; 10830/3; 10860/3; 10890/3; 10920/3; 10950/3; 10980/3; 11010/3; 11040/3; 11070/3; 11100/3; 11130/3; 11160/3; 11190/3; 11220/3; 11250/3; 11280/3; 11310/3; 11340/3; 11370/3; 11400/3; 11430/3; 11460/3; 11490/3; 11520/3; 11550/3; 11580/3; 11610/3; 11640/3; 11670/3; 11700/3; 11730/3; 11760/3; 11790/3; 11820/3; 11850/3; 11880/3; 11910/3; 11940/3; 11970/3; 12000/3; 12030/3; 12060/3; 12090/3; 12120/3; 12150/3; 12180/3; 12210/3; 12240/3; 12270/3; 12300/3; 12330/3; 12360/3; 12390/3; 12420/3; 12450/3; 12480/3; 12510/3; 12540/3; 12570/3; 12600/3; 12630/3; 12660/3; 12690/3; 12720/3; 12750/3; 12780/3; 12810/3; 12840/3; 12870/3; 12900/3; 12930/3; 12960/3; 12990/3; 13020/3; 13050/3; 13080/3; 13110/3; 13140/3; 13170/3; 13200/3; 13230/3; 13260/3; 13290/3; 13320/3; 13350/3; 13380/3; 13410/3; 13440/3; 13470/3; 13500/3; 13530/3; 13560/3; 13590/3; 13620/3; 13650/3; 13680/3; 13710/3; 13740/3; 13770/3; 13800/3; 13830/3; 13860/3; 13890/3; 13920/3; 13950/3; 13980/3; 14010/3; 14040/3; 14070/3; 14100/3; 14130/3; 14160/3; 14190/3; 14220/3; 14250/3; 14280/3; 14310/3; 14340/3; 1437

Gene: CESP:119D2.1
Introns: 56/3: 96/

Query Match	19.2%	Score 393;	DB 2;	Length 860;
Best Local Similarity	29.8%;	Pred. No. 5.2e-23;		
Matches 117; Conservative 54;	Mismatches 161;	Indels 60;	Gaps 19;	
y 8 VETLVVADESM---VKFHGADLEHYLTLLATAAR---- <td> : </td> <td> </td> <td> </td> <td> : </td>	:			:
b 135 VELAVFADDAWDHFKKMKYKAAEENMHTFIMAVVNNIDLVYQRLQOPRINKIVRYEI 194	:			:
y 60 LRD----RDSGPKVYTGNAALTLRNFCAMQKLNKVKSDRKHPEYWDTAIFLTRODLCGATTC 115	:			:
b 195 LKNIPHLNARKHSGVDVRLDLDAFCQYQNEINPPNDADPRHWDHALLFSGVDLHRNGVK 254	:			:
y 116 DTLGHADVGTMCDDPKRCSVIEDDGLPSAFTTAHELGHVFNPHDNVKVCEVFGKLRAN 175	:			:
b 255 TVAGVAPVKGMSGVKRSCTINEGLDFGSVVVTHEMGHSLGMYHGDGDECD-----LRC- 308	:			:
y 176 HMSPTLIQIDRANPWSACSAIIITFDLDSG-GD-----CLLD--OPSPKISLPE-D 224	:			:
b 309 CIMSPSV--GSGKTHWSQSVNEMATFV--GHLGDFFRPPNCLQDASANEQRVAFKGESE 364	:			:

```

y      |||||:||||:---ECWKHELKDGDTMNCIQMVWC---GNGEGVIRTAH-PA 416
o      |||||:||||:---ECWKHELKDGDTMNCIQMVWC---GNGEGVIRTAH-PA 416

```

[illegible]

RESULT 7
118856
 Caenorhabditis elegans inhibitor homolog - *Caenorhabditis elegans*
Species: *Caenorhabditis elegans*
Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
Accession: T18856; T24653
 MCMurray, A.
 Submitted to the EMBL Data Library, July 1995
 Reference number: Z19031
 Accession: T18856
 Status: preliminary; translated from GB/EMBL/DBDJ
 Molecule type: DNA
 Residues: 1-1444 <W1>
 Cross-references: EMBL: Z50004; PIDN: CAA90293.1; GSPDB: GN000028; CESP: C02B4.1
 Experimental source: clone C02B4

McMurray, A.
Submitted to the EMBL Data Library, July 1995
;Reference number: Z19917
;Accession: T24653

:Residues: 1-1444 <

;Cross-references: EMBL:J50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1
;Experimental source: clone T07C5
;Genetics:
;Gene: CESP:C02B4.1

A:Map position: X
A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3; 568

Query Match 18.5%; Score 377.5; DB 2; Length 1444;
Best Local Similarity 30.5%; Pred. No. 1.9e-21;
Matches 115; Conservative 45; Mismatches 170; Indels 47; Gaps 15;

QY 25 DLEHYL---LTLTAAARLYRHPSILNPIVIVVKKVLLLRDRS---GPKVTGNAALT 76
DB 239 DAEQHLLEFSLALNNHVLVQDQLTPNLDIVIVRYEMWRTQPSALSTGVHKNQQAQSL 298

QY 77 LRNFCAWQKLNKVS---KHPYVDTAILETRQDLGCGATTCTDGLMADVGMCDPKRSCSV 135
DB 299 LDAFCRYQAHNPNPGLTDMNHYDHGVLITGYDIYHTT---SVAGVAPVARMCDPLFACSL 357

QY 136 IEDDGLPSAFTAHGLHGVFNPHDNV---VCEEVFGKLRANHMSPILQIDRANPHSAC 194
DB 358 VEGHLGRSFLAHMGHNMGMVHDGVQNCCKCLMSAYNGAGKT-----TWSDC 409

QY 195 SAAIITDFL---DSGHGDCLLDQPSKPSLPE---DLPGASYTLQQCELAFAFGVGSK- 245
DB 410 SVREFNAFLQLDESGRNCGLRDLASPLISTNHLSDLPLQORTADQOCSYFWGRDYKV 469

QY 246 PCP---YMQYCTKLWCTGKAKGQMVQOTRHPFMDGTSCGEGKLCUKGACVE---RHNLNK 300
DB 470 EIPNGKAMDDICRILWCGNSG---TISTAH---PALEGSWCGANKWKHGQCTHTWTFGLTP 525

QY 301 HRVDSWAKW---PYGCSRTCCGGVOLARQCTNPTPANGGKYCEGVRYKYRSCN 354
DB 526 VPIDGSEWGAEGKCPICQCAVSGSITVQOQRDCVNPAPNNGGKTCEGANIRGIVCG 585

QY 355 LEPCCPSASGSKSFREQ 371
DB 586 ----ATSSNCLGFTREE 598

RESULT 8
T15976
hypothetical protein F08C6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15976
R:Bentley, D.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F08C6.
A:Reference number: 218440
A:Accession: T15976
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-957 <BEN>
A:Cross-references: EMBL:U29378; NID:g968184; PID:g968185; PIDN:AAA68721.1; CESP:F08C6.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F08C6.1
A:Introns: 23/1; 135/3; 220/3; 285/3; 325/2; 376/3; 461/2; 498/3; 532/3; 627/2; 714/1; 7

Query Match 14.6%; Score 297.5; DB 2; Length 957;
Best Local Similarity 25.2%; Pred. No. 1.9e-15;
Matches 116; Conservative 54; Mismatches 157; Indels 133; Gaps 19;

QY 8 VETLVVADSWKVF---HG--AD--LEHYLLTLTAAARLYRHPSILNPIVIVVKKVLL 59
DB 221 VELAVFVDENLWRHFSKGMADRKLQDYLTLNLIQIYQTPASPLFRVIREV 280

QY 60 LRDRDSG---PKVTGNAALTLRNFCAWQKLNKVS DKHPEYWDTAILETRQDL---CGATT 114
DB 281 LTRQPSALAGYLNHNGNAQWYLDRCRYQRLN---AVRD-----WDHAIMLTGYDIHRGAGS 334

QY 115 CDTLGMADVGMCDPKRSCSVIEDDGLPSAFTAHGLHGVFNPHDNVKNVCEVFGKLR 174
DB 335 RSISGIARLDGMCDPNTCTLABGLDFTSAFTGTHGLHRFTLKH-----KS 383

QY 175 NHMMSPTL-----IQIDRANPHSACS-----AAIIT-----DFLSDSHGDCLLD 213

DB 384 DTLETFGCGSKWQCGRCVPTWGTNETIQTVOHVAPVVTTLPSRIDGSMGSGWATICS 443
QY 214 Q-----PSKPISLPEDLPAS---YTLQOCELA----- 239

DB 444 QCTCNGILSGVGLAIARRTCSAPYPANGGSDCVGSTRAVLCSROGGRASKSVDEYISDK 503

QY 240 -----FGVSGKPCPYMOYCTKLWCTGKAKGQMVQOTRHPF---ADGTSC 281
DB 504 CMECKRLKNDRELTKGGSQNLNRFQORACKVFDVQ---QHYGSQRNRYRFGDNLPGDTSC 560

QY 282 GEKGLKLGACVHRNLNKHVD-----GSAWAKWDYPGCSRTC 320
DB 561 GYDEYCLDGECLALNCNNALISRDQSCPTDCTPTDQSSSVYRGWGTWSLWTSCTATC 620

QY 321 GGGVOLARQCTNPTPANGGKYCEGVRYKYRSCNLEPCPS 360
DB 621 GGGYRKNRACS---ITG---QCEGNEDETEVCSSSESCPS 654

RESULT 9
S60257
meltrin alpha - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
R:Yagami-Hironaka, T.; Sato, T.; Kurisaki, T.; Kamijo, K.; Nabeshima, Y.I.; Fujisawa-Nature 377, 652-656, 1995
A:Title: A metalloprotease-disintegrin participating in myoblast fusion.
A:Reference number: S60257; MUID:96026308; PMID:7566181
A:Accession: S60257
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-903 <YAG>
A:Cross-references: EMBL:DS0411; NID:g1054586; PIDN:BAA08912.1; PID:g1054587
C:Superfamily: mouse meltrin alpha; disintegrin homology
F:421-503/Domain: disintegrin homology <DIS>
F:349/Active site: Glu #status predicted

Query Match 11.4%; Score 233.5; DB 2; Length 903;
Best Local Similarity 21.1%; Pred. No. 1.9e-10;
Matches 107; Conservative 49; Mismatches 149; Indels 203; Gaps 16;

QY 2 VSIPIRYVETLVVADSWKVFHGADLEHYLLTLTAAARLYRHPSILNPIVIVVKKVLLLR 61
DB 208 LKMTKYVELVIVADNREFQGRGDLKVKORLEIANHVDKFFRPLN-IRIVLGVGVWN 266

QY 62 DRDGGPKVTGNAALTLRNFCAWQKLNKVS DKHPEYWDTAILETRQDLGCGATTCDTLGMA 121
DB 267 DIDR-CSIQSDPPTRLHEFLDWRKIKLLPRKSH---DNAQLISGVYFQGT---TIGMA 317

QY 122 DVTMCDPKRSCSVI---EDDGLPSAFTAHGLHGVFNPHDNV-KVCEEVFGKLRANHM 177
DB 318 PMSMCTAEQSGGVYMDHSDSPGAAVTLAHLGHNFHGMNHDTLERCCSRMAEKGGCI 377

QY 178 MSPILQIDRANP---WSACSAIITDFLDSHGDCLLDQP----- 215
DB 378 MNPS-----TGFPFPMVFSSCRKDLASLEKGMGMCLENLPEVKQAFGKRGKNGVVEE 432

QY 216 -----SKPISLPED 224
DB 433 GEEDCGPEBECTNRCCNATTCTLKPDVCAHGOCCEDCOLKPPGTACRGSSNSCDLPEF 492

QY 225 LPGAS-----YTLQOCELAFCVGSKPCPYMOY----- 252
DB 493 CTGTAPECPANVYLHDGHPGQGVGYNGICQTHEQQCVTLGWPGKAPAGTCFERNVS 552

QY 253 -----CTKLWCTGKAK-----GOMV 267
DB 553 AGDPYNGCGKDSKSAFAKACELRDAKCGKIQCGGASRPVIGTNAVSTETNIPQEGGRIL 612

QY 268 COTRHFPPWAD-----GTSCEGKCLKGACVHRNLNKHVDGWSWAKWDYPGCS 317
DB 317 -----GTSCEGKCLKGACVHRNLNKHVDGWSWAKWDYPGCS 317

[illegible]

QY 120 MADVTGMDPKRSCVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEEVFGKLRANHM 178
Db 278 -----SSVLELRLGLTIIVTTL-----QDSIRKVTTEE--NKELANELR 313
QY 179 SPTL-----IQIDRANPWSA-----CSAAIITDFLDSGHGD 209
Db 314 RPPLCYHNGVQYRNEEWTVDSCTECHCONSVTICKKVSCTPMPSCSNATVPD-----GE 367
QY 210 CLLDQPSKPSLSPED-----LPGASYT-----LSOOCELAFGVGS 244
Db 368 CC-----PRCWPSSDADGWSFWSEWTSCTSCGNGIQGRSCDLSNRCCE-GSSVQT 420
QY 245 KPCPYMOYCTKLWCTGKAKGMVQTRHFPWAD-GTSCGEG----- 284
Db 421 RTC-HIQECDKRF---KODGGW---SHWSPSSCVTCGDGVITRIRLNCNPSQPMNGKP 473
QY 285 -----KLCIKGACVERHNKRVHVDGWSAKWDPYGPCSRTCGGVLARQCTNPTP 336
Db 474 CEGEARETKACKKDAC-----PINGGWPSPWDICSVTCGGGVQKRSRLCNRNPTP 524
QY 337 ANGKGYCEGVYKVRSCNLEPCP 359
Db 525 QFGGKDCVGDVTENQICNKQDCP 547

RESULT 12
A40558
thrombospondin 1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 20-Aug-1999
C:Accession: A40558; B437905; B42587; S68787
R:Lawler, J.; Duquette, M.; Ferro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.
Genomics 11, 587-600, 1991
A:Title: Characterization of the murine thrombospondin gene.
A:Reference number: A40558; MUID:92128941; PMID:1774063
A:Accession: A40558
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1170 <LAH>
A:Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454; GB:M62455; GB:M62466; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:9511867; PIDN:AAA5061
R:Bornstein, P.; Alfli, D.; Devayayalu, S.; Franson, P.; Li, P.
J. Biol. Chem. 265, 16691-16698, 1990
A:Title: Characterization of the mouse thrombospondin gene and evaluation of the role of
A:Reference number: A37905; MUID:90375546; PMID:2398070
A:Accession: A37905
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <BOR>
A:Cross-references: GB:J05605; GB:J05606; NID:9201991; PIDN:AAA40431.1; PID:g554390
R:Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
A:Reference number: A42587; MUID:92147683; PMID:1371115
A:Accession: B42587
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1152, 'p', 1154-1170 <LAH>
A:Cross-references: GB:H87276
R:Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F.
FEBS Lett. 387, 36-41, 1996
A:Title: Expression and initial characterization of recombinant mouse thrombospondin 1 a
A:Reference number: S68787; MUID:96234006; PMID:8654563
A:Accession: S68787
A:Molecule type: protein
A:Residues: 19-26, 'X', 28-37 <CHE>
C:Complex: homotrimer, disulfide linked
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; v
C:Keywords: calcium binding; glycoprotein; homotrimer
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1170/Product: thrombospondin 1 #status predicted <MAT>
F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>

F:378-429/Domain: thrombospondin type 1 repeat homology <THR1>
F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
F:551-586/Domain: EGF homology <EGF>
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.08; Score 224.5; DB 2; Length 1170;
Best Local Similarity 24.3%; Pred. No. 1.3e-09;
Matches 93; Conservative 35; Mismatches 106; Indels 149; Gaps 20;

QY 63 RDSGPKVTGNAALTNFCAMQKLNKYSKHPVWDTAILFTRQDL---CGATTCDTLG 119
Db 228 RKGSSSTNILLTLDN-----NVNGSPAIRTYIGHKTKDLQAICG-LSCDEL- 277
QY 120 MADVTGMDPKRSCVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEEVFGKLRANHM 178
Db 278 -----SSVLELRLGLTIIVTTL-----QDSIRKVTTEE--NRRLVSELK 313
QY 179 SPTL-----IQIDRANPWSA-----CSAAIITDFLDSGHGD 209
Db 314 RPPLCYHNGVQYRNEEWTVDSCTECHCONSVTICKKVSCTPMPSCSNATVPD-----GE 367
QY 210 CLLDQPSKPSLSPED-----LPGASYT-----LSOOCELAFGVGS 244
Db 368 CC-----PRCWPSSDADGWSFWSEWTSCTSCGNGIQGRSCDLSNRCCE-GSSVQT 420
QY 245 KPCPYMOYCTKLWCTGKAKGMVQTRHFPWAD-GTSCGEG----- 284
Db 421 RTC-HIQECDKRF---KODGGW---SHWSPSSCVTCGDGVITRIRLNCNPSQPMNGKP 473
QY 285 -----KLCIKGACVERHNKRVHVDGWSAKWDPYGPCSRTCGGVLARQCTNPTP 336
Db 474 CEGEARETKACKKDAC-----PINGGWPSPWDICSVTCGGGVQKRSRLCNRNPTP 524
QY 337 ANGKGYCEGVYKVRSCNLEPCP 359
Db 525 QFGGKDCVGDVTENQICNKQDCP 547

RESULT 13
S24789
jararagin C precursor - jararaca (fragment)
N:Alternate names: single chain botrocetin
N:Contains: disintegrin-like 28k protein; hemorrhagic proteinase (EC 3.4.24.-)
C:Species: Bothrops jararaca (Jararaca)
C:Date: 20-Feb-1995 #sequence_revision 29-Aug-1997 #text_change 09-Jun-2000
C:Accession: S24789; JC2245; A44463; A37958; JC2373
R:Paine, M.J.I.
submitted to the EMBL Data Library, August 1992
A:Reference number: S24789
A:Accession: S24789
A:Molecule type: mRNA
A:Residues: 1-571 <PAI>
A:Cross-references: EMBL:X68251; NID:g62467; PID:g62468
R:Usami, Y.; Fujimura, Y.; Miura, S.; Shima, H.; Yoshida, E.; Yoshioka, A.; Hirano, K
Biochem. Biophys. Res. Commun. 201, 331-339, 1994
A:Title: A 28 kDa protein with disintegrin-like structure (jararagin-C) purified fro
A:Reference number: JC2245; MUID:94256999; PMID:8198592
A:Accession: JC2245
A:Molecule type: protein
A:Residues: 360-571 <USA>
A:Experimental source: venom
R:Paine, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.
J. Biol. Chem. 267, 22869-22876, 1992
A:Title: Purification, cloning, and molecular characterization of a high molecular we
ily.
A:Reference number: A44463; MUID:93054601; PMID:1385408
A:Accession: A44463
A:Molecule type: mRNA
A:Residues: 1-23, 'Q', 25-92, 'G', 94-131, 'G', 133-169, 'Q', 171-571 <PA>
A:Cross-references: GB:X68251
A:Experimental source: venom gland
A:Note: sequence inconsistent with the nucleotide translation

Fri May 2 10:41:46 2003

```
QY 176 HMMSPTLIQIDRANPWSAGSAAIITDFLDSHGDCILLDQP-----SKPI----- 219
Db 171 ISHEPSTF-----FSNCSYIQCWDFINDHNPFCIVNEPLGTDIVSPVCGNLELVEGE 223
QY 220 -----SLPEDLPGASYT 231
Db 224 EDCGTPENCNECCDAATCKLKSGSQGHGDCCEQCKFSKSGTECRESMPECDPAEHC 283
QY 232 -LSQOCEL-AFGVGSKPC-----P 248
Db 284 QSSSECPADVFKHNGQPCLDNTGYCYNGNCPIMYHOCYALWGADVYDAEDSCFESNTKGN 343
QY 249 YMOYCTK-----LMCTGKAKGQ-----MYCQTRHFPW 275
Db 344 YGYCRKENGKIPCAPEDVKCGRLYCKDNSPCQNNPCKMFYSNEDEHKGW----- 395
QY 276 ADGTSCEGKCLKGACVE 294
Db 396 LPGTKGCGDKVCSNGHCVD 414
```

Search completed: April 29, 2003, 17:13:21
Job time : 12.5813 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 4.56924 Seconds
(without alignments)
3367.676 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_583

Perfect score: 2042

Sequence: 1 FVSIPIRYVELVVADESWK.....SCNLEPCPSSASCKSFREQ 371

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1337	65.5	968	1 ATSL_MOUSE	P97857 mus musculus
2	1334	65.3	967	1 ATSL_HUMAN	Q9uh18 homo sapien
3	1327	65.0	967	1 ATSL_RAT	Q9wuq1 rattus norv
4	1182.5	57.9	837	1 ATSL_HUMAN	Q75173 homo sapien
5	1133.5	56.5	630	1 ATSL_RAT	Q9esp7 rattus norv
6	1128	55.2	905	1 ATSL_MOUSE	P57110 mus musculus
7	1126.5	55.2	890	1 ATSL_HUMAN	Q9up79 homo sapien
8	1114	54.6	930	1 ATSL_MOUSE	Q9r001 mus musculus
9	1110	54.4	930	1 ATSL_HUMAN	Q9una0 homo sapien
10	1046.5	51.2	1629	1 ATSL_HUMAN	Q9p2n4 homo sapien
11	819	40.1	997	1 ATSL_HUMAN	Q9ukp4 homo sapien
12	786	38.5	1593	1 ATSL_HUMAN	P58397 homo sapien
13	751	36.8	1077	1 ATSL_HUMAN	Q9h324 homo sapien
14	633.5	31.0	1211	1 ATSL_HUMAN	Q95450 h adamts-2
15	631.5	30.9	1205	1 ATSL_HUMAN	O15072 homo sapien
16	628.5	30.8	1205	1 ATSL_BOVIN	P79331 b adamts-2
17	620	30.4	860	1 ATSL_HUMAN	Q9ukp5 homo sapien
18	599	29.3	207	1 ATSL_BOVIN	Q9tt92 bos taurus
19	599	29.3	245	1 ATSL_BOVIN	Q9tt93 bos taurus
20	246.5	12.1	776	1 AD28_MACFA	Q9xl16 macaca fasc
21	236.5	11.5	909	1 AD12_HUMAN	O43184 macaca fasc
22	234	11.5	775	1 AD12_HUMAN	Q9ukq2 homo sapien
23	233.5	11.4	903	1 AD12_MOUSE	O61824 mus musculus
24	229.5	11.2	1170	1 TSP1_HUMAN	P07996 homo sapien
25	224.5	11.0	1170	1 TSP1_MOUSE	P35441 mus musculus
26	223.5	10.9	824	1 AD08_HUMAN	P78325 homo sapien
27	222.5	10.9	1170	1 TSP1_BOVIN	Q28178 bos taurus
28	220	10.8	571	1 DISJ_BOVJA	P30431 bothrops ja
29	219.5	10.7	774	1 AD28_MOUSE	Q9jln6 mus musculus
30	214	10.5	813	1 AD33_HUMAN	Q9bz11 homo sapien
31	210.5	10.3	920	1 AD19_MOUSE	O35674 mus musculus
32	210	10.3	416	1 HR1B_TRIFL	P20164 trimeresuru
33	205.5	10.1	936	1 AD19_HUMAN	Q9h013 homo sapien

RESULT 1

ID	ATSL_MOUSE	STANDARD;	PRT;	968 AA.
DT	P97857; 054768;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).			
GN	ADAMTS1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=129/SVJ;			
RX	MEDLINE=98110583; PubMed=9441751;			
RA	Kuno K., Lizasa H., Ohno S., Matsushima K.;			
RT	"The exon/intron organization and chromosomal mapping of the mouse			
RT	ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";			
RL	Genomics 46:466-471(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97150761; PubMed=8995297;			
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,			
RA	Matsushima K.;			
RT	"Molecular cloning of a gene encoding a new type of metalloproteinase-			
RT	disintegrin family protein with thrombospondin motifs as an			
RT	inflammation associated gene.";			
RL	J. Biol. Chem. 272:556-562(1997).			
RN	[3]			
RP	CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.			
RX	MEDLINE=99303657; PubMed=10373500;			
RA	Kuno K., Terashima Y., Matsushima K.;			
RT	"ADAMTS-1 is an active metalloproteinase associated with the			
RT	extracellular matrix.";			
RL	J. Biol. Chem. 274:18821-18826(1999).			
RN	[4]			
RP	FUNCTION.			
RX	MEDLINE=20389568; PubMed=10930576;			
RA	Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,			
RA	Ohno H., Matsushima K.;			
RT	"ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";			
RL	FEBS Lett. 478:241-245(2000).			
RN	[5]			
RP	FUNCTION, AND INDUCTION			
RX	MEDLINE=2043757; PubMed=10781075;			
RA	Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,			
RA	Richards J.S.;			
RT	"progesterone-regulated genes in the ovulation process: ADAMTS-1 and			
RT	cathepsin L proteases.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).			
CC	-!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE			
CC	INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY			
CC	SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH			

ALIGNMENTS

34	199	9.7	826	1	AD08_MOUSE	Q05910 mus musculus
35	197	9.6	413	1	ACIA_AKGAC	Q9pw35 agkistrodon
36	197	9.6	480	1	DISA_TRIGA	P15503 trimeresuru
37	197	9.6	797	1	AD33_MOUSE	Q923w9 mus musculus
38	196.5	9.6	814	1	AD15_HUMAN	Q13444 homo sapien
39	194	9.5	1074	1	SMSA_HUMAN	Q13591 homo sapien
40	192.5	9.4	1093	1	SMSB_MOUSE	Q60519 mus musculus
41	187.5	9.2	867	1	SSPO_BOVIN	P98167 bos taurus
42	186.5	9.1	1077	1	SMSA_MOUSE	Q62217 mus musculus
43	184	9.0	1173	1	TSP1_XENLA	P35448 xenopus lae
44	183.5	9.0	203	1	FIBR_AGKCO	P28891 agkistrodon
45	181.5	8.9	414	1	HRTD_CROAT	P15167 crotalus at

CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-1-LEU-1692
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX.
CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC CELLS OF PREOVULATORY FOLLICLES.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001735; BAA24501.1; ALT INIT.
CC EMBL; D67076; BAA11088.1; ALT_FRAME.
CC MEROPS; M12.222; -.
CC MGD; MGI:109249; Adamt1.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_prop.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; zn_Mtpeptidse.
CC Pfam; PF00090; tsp_1; 3.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_prop; 1.
CC SMART; SM00209; TSP1; 3.
CC PROSITE; PS02115; ADAM_MEPRO; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS00092; TSP1; 3.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix; Heparin-binding.
CC SIGNAL 1 48
CC PROPEP 49 253
CC CHAIN 254 968
CC SITE 206 206
CC METAL 402 402
CC METAL 403 403
CC ACT_SITE 406 406
CC METAL 412 412
CC METAL 477 559
CC DOMAIN 560 617
CC DOMAIN 618 725
CC DOMAIN 726 850
CC DOMAIN 851 909
CC DOMAIN 910 968
CC DOMAIN 195 199
CC DOMAIN 548 548
CC CARBOHYD 721 721
CC CARBOHYD 765 765
CC CARBOHYD 783 783
CC CARBOHYD 946 946
CC MUTAGEN 403 403
CC CONFLICT 335 335
CC CONFLICT 425 425
CC SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;
SQ

Query Match 65.5%; Score 1337; DB 1; Length 968;
Best Local Similarity 63.2%; Pred. No. 2e-103;
Matches 237; Conservative 51; Mismatches 81; Indels 6; Gaps 4;
QY 1 FVSPRYVETLVVADESNVWFHAGDALEHYLLTLLATAARLYRHPISILNPINIVVVKVLL 60
Db 254 FVSPRYVETLVVADESNVWFHAGDALEHYLLTLLATAARLYRHPISILNPINIVVVKVLL 313
QY 61 RDRSDGPKVTGNAALTRNFCANQKLNKVKYDKHPEYWDTAIFLTRDLCGATTCOTLGM 120
Db 314 YEEQKGPEVTSNAALTRNFCANQKLNKVKYDKHPEYWDTAIFLTRDLCGATTCOTLGM 373
QY 121 ADVCTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFKLRANHMSP 180
Db 374 ADVCTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFKLRANHMSP 433
QY 181 TLIQIDRANPWSACSAIITDFLDHSGHDCGLDQPSKPISLPDLPGASTVLSQOCELAF 240
Db 434 MLSLIDHSQWSPSCSAIITDFLDHSGHDCGLDQPSKPISLPDLPGASTVLSQOCELAF 493
QY 241 GVGSKPCP-YMQYCTKLWCTGKAKGMVQCOTRHPFWADGTSCEGKCLKAGACVVERHNLN 299
Db 494 GEESKHCPDAASTCTTLCCTGSGLLVCQTKHFPWADGTSCEGKCLKAGACVVERHNLN 552
QY 300 KH--RVDSWAKWDPYGPCSRTCCGGVQLARRCTNPTPANGGKYCEGVRYRSCNLE 356
Db 553 KHFPATVHGSGWPGPWPNGDCSRTCCGGVQVYTMRECDNPVPNGSKYCEGKRVYRSCNIE 612
QY 357 PCPSASGSKSPREQ 371
Db 613 DCPDN-NGKTFREQ 626
RESULT 2
ATSL_HUMAN STANDARD; Q9UH83; Q9P2K0; Q9NSJ8;
ID ATSL_HUMAN PRT; 967 AA.
AC Q9UH83; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
GN ADAMTS1 OR METH1 OR KIAA1346.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
RT "Cloning, characterization and mapping on human chromosome 21 of the
RL orthologue of murine Adamts-1.";
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RA Tissue=Heart;
RC MEDLINE=93367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RL family of proteins with angio-inhibitory activity.";
RN [3]
RP SEQUENCE FROM N.A.
RA Tissue=Endothelial cells;
RC MEDLINE=20247184; PubMed=10785405;
RA Glienke J., Schmitt A.O., Filarsky C., Hinzmann B., Weiss B.,
RA Rosenthal A., Thierach K.H.;
RT "Differential gene expression by endothelial cells in distinct
RL angiogenic states.";
RN [4]
RP Eur. J. Biochem. 267:2820-2830(2000).
SQ SEQUENCE FROM N.A.

RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakai Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordtsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Reinhardt R., Yaspou M.-L.,
RA Lehrach H., Reinhardt R., Yaspou M.-L., Nizetic D., Francis F.,
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [6]
RP SEQUENCE OF 418-967 FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=20289799; PubMed=10830953;
RA Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR
CC ACTIVITY. ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.
CC -!- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1938-GLU-| LEU-1939
CC SITE. WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is not modified and
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF170084; AAF15317.1;
CC EMBL: AF060152; AAD48080.1; ALT_INIT.
CC EMBL: AF207664; AAF23772.1;
CC EMBL: AB037767; BAA92584.1; ALT_INIT.
CC EMBL: AF001697; BAA95502.1;
CC EMBL: AL162080; CAB82413.1;
CC MEROPS: M12.222;
CC Genew: HGNC:217; ADAMTS1.
CC MIM: 605174;
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Repolysin.
CC InterPro: IPR000884; TSP1.
CC InterPro: IPR000130; Zn_MTPeptase.
CC Pfam: PF00090; tsp.1; 6.
CC Pfam: PF01421; Repolysin; 2.
CC Pfam: PF01562; Pep_M12B_propep; 2.
CC SMART: SM00209; TSP1; 3.

DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 49
FT PROPEP 50 252
FT CHAIN 253 967
FT SITE 198 198
FT METAL 401 401
FT ACT_SITE 402 402
FT METAL 405 405
FT METAL 411 411
FT DOMAIN 476 559
FT DOMAIN 560 616
FT DOMAIN 617 724
FT DOMAIN 725 849
FT DOMAIN 850 908
FT DOMAIN 909 967
FT DOMAIN 843 846
FT CARBOHYD 547 547
FT CARBOHYD 720 720
FT CARBOHYD 764 764
FT CONFLICT 227 227
FT CONFLICT 468 468
FT CONFLICT 561 561
FT SEQUENCE 967 AA; 105383 MW; C189389324741ED1 CRC64;
Query Match 65.3%; Score 1334; DB 1; Length 967;
Best Local Similarity 63.5%; Pred No. 3.5e-103;
Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;
QY 1 FVSIPIRYVETLVVADSMVKFHHGADLEHYLLTLAARLYRHPSLINP INIVVVKVLL 60
Db ||| |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 60
Db 253 FVSSHYRYVETLVVADSMVKFHHGADLEHYLLTLAARLYRHPSLINP INIVVVKVLL 312
QY 61 RDRDSGPKVTGNAALTLRNFCWQKLNKVDKHPYWDYDTALFTRODLGGATCTDGLM 120
Db ||| |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 120
Db 313 HDEQKGEVTSNAAALTLRNFCWQKLNKVDKHPYWDYDTALFTRODLGGATCTDGLM 372
QY 121 ADVTGMDCKRSCSVIEDDGLPSAFTTAHELGHVFNHMDNVKVCVEEFGKLRANHMS 180
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 180
Db 373 ADVTGMDCKRSCSVIEDDGLPSAFTTAHELGHVFNHMDNVKVCVEEFGKLRANHMS 432
QY 181 TLIIQIDRANPWSACAAITITFLDSHGDCDLDQSPKISLPEDLPASVYLSQCELA 240
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 240
Db 433 MLSNLDHSQPSWPCSAAYMITSLFNGHGECLMDKPNPQLPGDLPSTYDANROCOPT 492
QY 241 GVGSKPCP-YMOYCTKLWCTGKAGOMVCOYTRHFWADTSGEGKGLKGCACVERHNL 299
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 299
Db 493 GEDSKHCPDAASTCTLWCTGTSGGVLCQTKHFWADTSGEGKGLKGCACVERHNL 551
QY 300 KH---RVGDSMAKWPDPYPCSRCTCGGVOLARRQCTNPANGKYCYGVRKYRSCNLE 356
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 356
Db 552 KHFDTPFHGSMWGMGPGWGDSCRTCGGVQYTMRECDNPVKNPKNGKYCYGVRKYRSCNLE 611
QY 357 PCPSSASGSKSPREQ 371
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 611
Db 612 DCPDN-NGKTPREQ 625
RESULT 3
ATSL_RAT
ID ATSL_RAT STANDARD; PRT; 967 AA.
AC Q9WUQ1; Q9ER11;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
 RA Little S.P.;
 RT "Induction of a disintegrin and metalloprotease with the
 RT thrombospondin type I motif (ADAMTS).";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 18-967 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=20304099; PubMed=10847486;
 RA Diamantis I., Luechi M., Hoesli M., Reichen J.;
 RT "Cloning of the rat ADAMTS-1 gene and its down regulation in
 RT endothelial cells in cirrhotic rats.";
 RL Liver 20:165-172(2000).
 CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
 CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
 CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
 CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684
 CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (by similarity).
 CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
 CC CIRRHOTIC LIVER.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF149118; AAD34012.1; -;
 DR EMBL; AF304446; AAG29823.1; -;
 DR MEROPS; M12.222; -;
 DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR002870; Pep_M12B_propep.
 DR InterPro; IPR001590; Reprolysin.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR000130; Zn_M12Bptdase.
 DR Pfam; PF00090; tsp_1; 6.
 DR Pfam; PF01421; Reprolysin; 2.
 DR Pfam; PF01562; Pep_M12B_propep; 2.
 DR SMART; SM00209; TSP1; 3.
 DR PROSITE; PS50215; ADAM_MEPRO; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS50092; TSP1; 2.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Hydrolase; Metalloprotease; Zinc; Signal; Heparin-binding.
 KW Repeat; Extracellular matrix; Reprolysin-binding.
 FT SIGNAL 1 54
 FT PROPEP 55 252
 FT CHAIN 253 967
 FT SITE 205 205
 FT METAL 401 401
 FT ACT_SITE 402 402
 FT METAL 405 405
 FT ZINC (CATALYTIC) (BY SIMILARITY).
 FT Cysteine switch (POTENTIAL).
 FT ZINC (CATALYTIC) (BY SIMILARITY).
 FT ZINC (CATALYTIC) (BY SIMILARITY).
 FT ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 411
 FT DOMAIN 476
 FT DOMAIN 559
 FT DOMAIN 616
 FT DOMAIN 724
 FT DOMAIN 725
 FT DOMAIN 858
 FT DOMAIN 908
 FT DOMAIN 914
 FT DOMAIN 919
 FT CARBOHYD 547
 FT CARBOHYD 720
 FT CARBOHYD 764
 FT CARBOHYD 782
 FT CARBOHYD 945
 FT CARBOHYD 945
 FT CONFLICT 21
 FT CONFLICT 26
 FT CONFLICT 49
 FT CONFLICT 72
 FT CONFLICT 79
 FT CONFLICT 249
 FT CONFLICT 262
 FT CONFLICT 607
 FT CONFLICT 936
 FT CONFLICT 962
 SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCB4CF CRC64;
 Query Match 65.0%; Score 1327; DB 1; Length 967;
 Best Local Similarity 62.7%; Pred. No. 1.3e-102;
 Matches 235; Conservative 52; Mismatches 82; Indels 6; Gaps 4;
 QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLL 60
 DB 253 FVSSPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLL 312
 QY 61 RDRSGKVTGNALTLNFCQWQKLNKSKHPYWDYLLTRDLCGATTCDTLGM 120
 DB 313 YEEQKGEVTSNAALTLNFCQWQKLNKSKHPYWDYLLTRDLCGATTCDTLGM 372
 QY 121 ADVGTMCDPKRSCVIEDDGLPSAFTTAHGLHGVNMPHDNPKVCEVEFGKLRANHMSP 180
 DB 373 ADVGTMCDPKRSCVIEDDGLPSAFTTAHGLHGVNMPHDNPKVCEVEFGKLRANHMSP 432
 QY 181 TLIIQIDRANPSACSAIITFLDSHGDCDLDQSPKISLPEDLPAGSYLSQCELA 240
 DB 433 MLSSLDHSPWSPCSAYMVTSPDLNKGHCBLMDKPNPKLPDLPGTLYDANRQCQTF 492
 QY 241 GVGSRPCP-YMOYCYKLCWCTGAKQWYVOTRHFWDGTSCGEGKLCIGKACVHRHNL 299
 DB 493 GEESTHCDDAATCTSTLWCTGTSGGLLVQCTKHFPWADGTSCGEGKWCYSGKVNKTDH- 551
 QY 300 KH--RVDGSKAWKDPYPCSRCTCGGVQVQLARRQCTNPTPANGKYCEGVRYKRSCHLE 356
 DB 552 KHFPATPVHSGWPGWPGWGDSCRTCGGVQVYTMRECDNPVPKNGKYCEGKRVYRSCNIE 611
 QY 357 PCPSSASGSKSFREQ 371
 DB 612 DCPDN-NGKTFREQ 625
 RESULT 4
 ID ATSA_HUMAN STANDARD; PRT; 837 AA.
 AC 075173; Q9UN83;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
 DE (ADMP-1).
 GN ADAMTS4 OR KIAA0688.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

RN RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE-98403880; PubMed-9734811;
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
 RA Kotani H., Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:169-176(1998).
 RN [2]
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE-99286303; PubMed-10356395;
 RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
 RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
 RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
 RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
 RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
 RA Trzaskos J.M., Arner E.C.;
 RA "Purification and cloning of aggrecanase-1: a member of the ADAMTS
 RT family of proteases.";
 RL Science 284:1664-1666(1999).
 RN [3]
 RN RP SEQUENCE FROM N.A.
 RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;
 RA "ADAMTS-4 genomic locus.";
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RX MEDLINE-20400518; PubMed-10827174;
 RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,
 RA Burn T.C., Arner E.C.;
 RA "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
 RT aggrecan substrate recognition and cleavage.";
 RL J. Biol. Chem. 275:25791-25797(2000).
 CC -1- INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
 CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
 CC ALZHEIMER'S DISEASE.
 CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-1-Ala-393
 CC site.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: secreted. Associated with the extracellular
 CC matrix (by similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
 CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
 CC -1- INDUCTION: BY INTERLEUKIN-1.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB014588; BAA31631.1; -;
 CC EMBL; AF148213; AAD41494.1; -;
 CC EMBL; AY044847; AAL02262.1; -;
 CC MEROPS; M12.221; -;
 CC Genew; HGNC:220; ADAMTS4.
 CC MIM; 603876; -;
 CC InterPro; IPR001762; Disintegrin.
 CC InterPro; IPR001590; Reprolysin.
 CC InterPro; IPR000884; TSP1.

DR InterPro; IPR000130; zn_MTpeptdse.
 DR Pfam; PF00090; tsp_1; 1.
 DR SMART; SM01421; Reprolysin; 1.
 DR SMART; SM00209; TSP1; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS00215; ADAM_MEPRO; 1.
 DR PROSITE; PS00092; TSP1; 1.
 DR PROSITE; PS00427; DISINTEGRIN-1; FALSE_NEG.
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Extracellular matrix.
 FT SIGNAL 1 51 POTENTIAL.
 FT PROPEP 52 212
 FT CHAIN 213 837
 FT SITE 194 194
 FT METAL 361 361
 FT ACT_SITE 362 362
 FT METAL 365 365
 FT METAL 371 371
 FT DOMAIN 437 519
 FT DOMAIN 520 576
 FT DOMAIN 577 685
 FT DOMAIN 686 837
 FT DOMAIN 247 252
 FT CARBOHYD 68 68
 FT CONFLICT 77 77
 FT CONFLICT 626 626
 FT CONFLICT 682 682
 FT CONFLICT G -> R (IN REF. 3).
 SQ SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;
 Query Match 57.9%; Score 1182.5; DB 1; Length 837;
 Best Local Similarity 55.5%; Pred. No. 1.2e-90;
 Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;
 Qy 1 FVSIPRIVETILVADESWKFEHGNADLEHYLLTLTAARLYRHPSTLNINIVVKKVLL 60
 Db 213 FASLSRFVETILVADDDKMAAFHAGLAKRYLLTVMAAAARAFKHPSTLRNPVSLVTRLVIL 272
 Qy 61 RDRDSGPKVTGNAALTIRNECAMOKLNKYSDKHPEYWDATLFTFRQDLGCGATTCDTLGM 120
 Db 273 GSGEQPGQVPSAAQTLRSFCAMQRLTPEDSDPDFDFTAILFTFRQDLGCGVSTCDTGM 332
 Qy 121 ADVGTMCDPKSCSVIEDGLPSAFTTAHELGHVFNPHDNVKNVCEVEKGL-RANHMS 179
 Db 333 ADVGTVCDPARSAIVEDDGLQSAFTAAHELGHVFNHMDNSRPSISLNGPLSTSRHWA 392
 Qy 180 PTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQPSKIPSLPDLFGASYTILSOCELA 239
 Db 393 PVMAHVDPPEPSPCSARFITDFLDNGYGHCLLDKPEAPLHPVTPFGKYDADRQCQLT 452
 Qy 240 FGVSCKPCPYM-QYCTKLWCTGKAKGMVCTRHFPWADTSCGEGKLCGLKACVERHNL 298
 Db 453 FGPOSRRHCPQLPPPCALWCSGHLNGHAMCOTKHSFWADGTGCGPAQACGGHCLHMDQL 512
 Qy 299 NKHRVD--GSNAKWDYPGCSRTCGGVQVQLARROCTNPTPANGKCYGCVYRKYRSCNLE 356
 Db 513 QDFNIPQAGGWGPGWGCDSRTCTCGGVQVFSRRDCTRPVPRNGKCYGEGRRTRFSCNTE 572
 Qy 357 PCPSASAKSPREEQ 371
 Db 573 DCP-TGSALTTPREEQ 586
 RESULT 5
 ATSA_RAT STANDARD; PRT; 630 AA.
 AC Q9ESP7; Q9ESP8; Q9ESP6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
 DE (Fragment).
 GN ADAMTS4.


```

FT ACT_SITE 365 365 BY SIMILARITY.
FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 439 526 DISINTEGRIN-LIKE.
FT DOMAIN 527 583 TSP TYPE-1 1.
FT DOMAIN 584 690 CYS-RICH.
FT DOMAIN 691 832 SPACER.
FT DOMAIN 833 890 TSP TYPE-1 2.
FT DOMAIN 890 930 POLY-PRO.
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 600 600 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 195 195 E -> R (IN REF. 2).
FT CONFLICT 413 440 YFTELLDGGHGGCLDADGALPLPTGL -> FSGCHLQGW
FT SEQUENCE 890 AA; 96671 MW; 57D70E03D5739D3 CRC64;

Query Match 55.2%; Score 1126.5; DB 1; Length 890;
Best Local Similarity 53.3%; Pred. No. 5.6e-86;
Matches 203; Conservative 64; Mismatches 101; Indels 13; Gaps 7;

QY 1 FVSIPRYVETLVVADSMVKFHGADLEHYLLTLLATAARLYRHPSTLNINIVVVKVLL 60
DB 215 FVSEARFVETLVVADSMVAFYGADJONHILTLMSVAARIYKHPSTLNINIVVVKVLL 274
QY 61 RDRSDGPKVGNALTLRNFCAWKLNKYSKHPYWDATLFTQDLCGAT-TCDDTLG 119
DB 275 EDEKNGPEVSDNGLTLRNFCAWKLNKYSKHPYWDATLFTQDLCGAT-TCDDTLG 334
QY 120 MADVTGMDPKRSCVIEDDGLPSATTAHELGHVFNMDHNVKCEEVFGKLRANHMS 179
DB 335 VADITGTCDPKSCVIEDDGLPSATTAHELGHVFNMDHNVKCEEVFGKLRANHMS 394
QY 180 PTLQIDRANPWSNCAAIITDFLDGSHGCLLDQPSKPLSLPEDLPG--ASYTLSQOCE 237
DB 395 PLFVHLNQTLPWSPCSAMYTELDDGGHGGCLLDADGALPLTGLPGRMALYQLDQOQR 454
QY 238 LAFVGSGKPCP---YMOYCTKLMCTGKAKGMVCOYTRH--FPWADGTCGEGKCLKLGAC 292
DB 455 QIFGPDPRHCPNTSAQDVCAQLWCHTDG-AEPLCHTKNGSLPWADGTCGEGKCLKLGAC 513
QY 293 VERHNLNKR--VDGSHAKWDPGCSRTCCGGVQVLAARCTNPTANGKCYEGVRVY 350
DB 514 LPEEVEERPKPVVDGWPAPGPGECRTCCGGVQVLAARCTNPTANGKCYEGVRVY 573
QY 351 RSCNLEPCPSSAGSKSPREQ 371
DB 574 QSCHEBPCP--PDGKSPREQ 592

RESULT 8
ID AT55_MOUSE STANDARD; PRT; 930 AA.
AC Q8R001;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (Implantin).
GN ADAMTS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.L., Hirohata S., Seidlin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT zinc metalloproteinases.";
```

```

J. Biol. Chem. 274:25555-25563(1999).
-|- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
-|- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-1-Ala-393 site.
-|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
-|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).
-|- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR UNDETECTABLE LEVEL THEREAFTER.
-|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
-|- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY SIMILARITY).
-|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
-|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
-|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).

EMBL; AF140673; AAD56356.1; -.
MEROPS; M12.225; -.
MGD; MGI:1346321; Adamts5.
InterPro; IPR001762; Disintegrin.
InterPro; IPR002870; Pep_M12B_propep.
InterPro; IPR001590; Reprolysin.
InterPro; IPR000884; TSP1.
InterPro; IPR000130; Zn_Mtpeptidse.
Pfam; PF00090; Tsp_1; 2.
Pfam; PF01421; Reprolysin; 1.
Pfam; PF01562; Pep_M12B_propep; 1.
SMART; SM00209; TSP1; 2.
PROSITE; PS00215; ADAM_MEPRO; 1.
PROSITE; PS00142; ZINC_PROTEASE; 1.
PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 261 POTENTIAL.
FT CHAIN 262 930 ADAMTS-5.
FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 411 411 BY SIMILARITY.
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 566 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE-1 1.
FT DOMAIN 624 731 CYS-RICH.
FT DOMAIN 732 874 SPACER.
FT DOMAIN 875 930 TSP TYPE-1 2.
FT DOMAIN 930 930 POLY-ALA.
FT DOMAIN 930 930 POLY-ARG.
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 930 AA; 101780 MW; 84DE84B26170D4DC CRC64;
```

Query Match 54.6%; Score 1114; DB 1; Length 930;
Best Local Similarity 53.6%; Pred. No. 6.4e-85;
Matches 201; Conservative 58; Mismatches 106; Indels 10; Gaps 3;

QY 2 VSPRYVETLVVADSMVKFHGADLEHYLLTLLATAARLYRHPSTLNINIVVVKVLLR 61

```
Db 263 ISRAQVELLVADSSMARMYGRGLQHYLLTLASIANFLYSHASIEHRIAVKVVVLT 322
QY 62 DRDSPKVTGNAULTLRNFCWQKLNKVKSPKPEYWDTAILFTQDLCCGATTCDTLGM 121
Db 323 DKDTSLVYSKNAATTLKFKCQWQHNGHNGDDEHYDAAALFTREDLGHSCDTLGM 382
QY 122 DVGTCMDPKRSCSVIEDGLPSAFTTAHELGLHVNPHDNVYKVEEYFGKLRNHHMSPT 181
Db 383 DVGTCSPERCANVEDDGLHAFTVAHEIGHLLGLSHDSKFCFEENFGTTERKLMSI 442
QY 182 LIQDRANPWSACAAITFDLSDHGDCLLDQSPKISLPEDLPGASYTLISQOCELAF 241
Db 443 LTSIDASKPWSKTSATITEFLDDHGNCLLDLPKQLGLPEELPGQTYDQOCNLTFG 502
QY 242 VGSRCPCYMYCTKLWCTGKAGQWCVOTRFPWADGTSCGKLCILKACVER-----H 296
Db 503 PEYSVCPGMDVCAWLWCVAVRGQWVCLTKLPKPAVEGTCGKRVCLQKCVDRKKYY 562
QY 297 NLNKHVRDGSWAKWDPYPCSRCTCGGVQVLARRCTNPTPANGKCYCEGVKVRKNSCLE 356
Db 563 STSSH---GNWGSNCPWGQCSRGCGGVQVFAIRHCNPNPANSRYCTGKRAIYRSCVT 619
QY 357 PCPSSASGSPREQ 371
Db 620 PCP--PNKGSFRHEQ 632

RESULT 9
ID AT55_HUMAN STANDARD; PRT; 930 AA.
AC Q9UNAO; Q9URP2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with chromospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (ADAM-TS 11).
GN ADAMTS5 OR ADMP2 OR ADAMTS11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99367476; PubMed=10438522;
RA Abbaszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H.,
RA Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M.,
RA Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K.,
RA Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H.,
RA Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F.,
RA Arner E.C., Burn T.C.;
RA "Cloning and characterization of ADAMTS11, an aggrecanase from the
RA ADAMTS family.";
RL J. Biol. Chem. 274:23443-23450(1999).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Kumf K., Lehmann R., Patterson D.,
RA Menzel U., Delabar J., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal K., Rump A., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Scharie M., Schoen O., Nordsiek G., Hornischer K., Brandt P.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
```

```
RN [3]
RP SEQUENCE OF 413-930 FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT zinc metalloproteases.";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
CC PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-1-Ala-393
CC site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
CC BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
CC CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,
CC CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
CC ARTHRITIC PATIENT.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
DR EMBL; AF142099; AAD49577.1; -
DR EMBL; AF001698; BAA95504.1; -
DR EMBL; AF001697; BAA95503.1; -
DR EMBL; AF141293; AAF02493.1; -
DR HSSP; Q9PW35; 1BUD.
DR MEROPS; M12.225; -.
DR Genew; HGNC:221; ADAMTS5.
DR MIM; 605007; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS50092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
FT SIGNAL 1 16
FT PROPEP 17 261 POTENTIAL.
FT CHAIN 262 930 ADAMTS-5.
FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 566 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE-1 1.
FT DOMAIN 624 731 CYS-RICH.
FT DOMAIN 732 874 SPACER.
```

```
FT DOMAIN 875 930 TSP TYPE-1 2.
FT DOMAIN 37 41 POLY-ALA.
FT DOMAIN 257 261 POLY-ARG.
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 138 138 A -> G (IN REF. 2).
FT CONFLICT 614 614 R -> H (IN REF. 3).
FT CONFLICT 692 692 P -> L (IN REF. 2).
FT SEQUENCE 930 AA; 101715 MW; B64281502F28193B CRC64;

Query Match 54.4%; Score 1110; DB 1; Length 930;
Best Local Similarity 53.6%; Pred. No. 1.4e-84;
Matches 201; Conservative 56; Mismatches 108; Indels 10; Gaps 3;

QY 2 VSPRYVETLVVADESVMKPHGADLEHYLLTLTAARLYRHPHSIINPINIVVVKVLLLR 61
DQ 263 ISRAQVELLVADASMARLYGRGLQHYLLTLASIANRLYSHASIEHRLAVRVVVLG 322
QY 62 DRSGPKVTGNAALTIRNFCWOKKLNKYSKHPYWDYTAIFTRDLCGAVTCDTLGMA 121
DQ 323 DKXSLVSNAAATLTKFKCKWQHQNGLGDDHEHYDAALFTREDLCGHHSDFLMA 382
QY 122 DVGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNKVKCEVEVFKLRANHHMSPT 181
DQ 383 DVGITCSPERSCAVIEDDGLHAAFTVAHEIGLHGLSHDSDKFCETFGSTEDKRLMSSI 442
QY 182 LIQIDRANPHSACSAIITDFLDSGHDCCLLDQSPKISIPEDLPASVTLSSOCELARG 241
DQ 443 LTSIDASKPKWSKTSATITEFDLDDGHGNCCLLDLPKQILGPEELPGQTYDATQCCNLTFG 502
QY 242 VGSKPCPYMOCYKTLGAKQGMVQCTRHFPWADTSCGEGKCLKLGACVER-----H 296
DQ 503 PEYSVCFGMVDCARLWCAVYRQGMVCLTKKLPAVESTPGCGKRICLQKGCVDKTKKVV 562
QY 297 NLNKHVRDGSWAKWDYPCSPRCGGVQLARQCTNPTPANGKCKYCEGVYKRYSCNLE 356
DQ 563 STSSH---GNWGSNGSGQCSRSRGGVQFAYRHNNPAPRNNRGYCTGKRAIYRSCSLM 619
QY 357 PCPSASGKGFREQ 371
DQ 620 PCP--PNGKSERHEQ 632

RESULT 10
AT9_HUMAN STANDARD; PRT; 1629 AA.
ID Q9P2N4; Q9NR29;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
GN ADAMTS9 OR KIAA1312.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Petal;
RX MEDLINE=203961138; PubMed=10936055;
RA Clark M.E., Kellner G.S., Turbeville L.A., Boyer A., Arden K.A.,
RA Maki R.A.;
RT "ADAMTS 9, a novel member of the ADAM-TS/Metallospandin gene
family.";
RL Genomics 67:343-350(2000).
RN [2]
RP SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
```

```
RT RT
RT RT
RL DNA Res. 7:65-73(2000).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.
CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,
CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN
CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR
CC THYMUS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF261918; AAF89106.1; -.
DR EMBL; AB037733; BAA92550.1; -.
DR HSP; P15167; IATL.
DR MEROPS; M12.021; -.
DR Genew; HGNC:13202; ADAMTS9.
DR MIM; 605421; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF00090; tsp_1; 11.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 12.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS50092; TSP1; 9.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 18
FT PROPEP 19 287
FT CHAIN 288 1629
FT DOMAIN 509 587
FT DOMAIN 589 642
FT DOMAIN 643 752
FT DOMAIN 753 880
FT DOMAIN 999 1053
FT DOMAIN 1056 1108
FT DOMAIN 1111 1156
FT DOMAIN 1184 1239
FT DOMAIN 1240 1295
FT DOMAIN 1332 1383
FT DOMAIN 1386 1439
FT DOMAIN 1445 1498
FT DOMAIN 1501 1554
FT DOMAIN 1562 1612
FT DOMAIN 88 96
FT SITE 223 223
FT METAL 434 434
FT ACT_SITE 435 435
```

*Prediction of the coding sequences of unidentified human genes. XVI.
The complete sequences of 150 new cDNA clones from brain which code
for large proteins in vitro.*;
DNA Res. 7:65-73(2000).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.
CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,
CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN
CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR
CC THYMUS.

CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
SIMILARITY).

CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
SIMILARITY).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.

CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AF261918; AAF89106.1; -.
DR EMBL; AB037733; BAA92550.1; -.
DR HSP; P15167; IATL.

DR MEROPS; M12.021; -.
DR Genew; HGNC:13202; ADAMTS9.

DR MIM; 605421; -.
DR InterPro; IPR001762; Disintegrin.

DR InterPro; IPR002870; Pep_M12B_propep.

DR InterPro; IPR001590; Reprolysin.

DR InterPro; IPR000884; TSP1.

DR InterPro; IPR000130; Zn_Mtpeptidse.

DR Pfam; PF00090; tsp_1; 11.

DR Pfam; PF01421; Reprolysin; 1.

DR Pfam; PF01562; Pep_M12B_propep; 1.

DR SMART; SM00209; TSP1; 12.

DR PROSITE; PS00215; ADAM_MEPRO; 1.

DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.

DR PROSITE; PS50092; TSP1; 9.

DR PROSITE; PS00142; ZINC_PROTEASE; 1.

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Alternative splicing.

FT SIGNAL 1 18
FT PROPEP 19 287
FT CHAIN 288 1629
FT DOMAIN 509 587
FT DOMAIN 589 642
FT DOMAIN 643 752
FT DOMAIN 753 880
FT DOMAIN 999 1053
FT DOMAIN 1056 1108
FT DOMAIN 1111 1156
FT DOMAIN 1184 1239
FT DOMAIN 1240 1295
FT DOMAIN 1332 1383
FT DOMAIN 1386 1439
FT DOMAIN 1445 1498
FT DOMAIN 1501 1554
FT DOMAIN 1562 1612
FT DOMAIN 88 96
FT SITE 223 223
FT METAL 434 434
FT ACT_SITE 435 435

```

FT METAL 438 438 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 444 444 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 749 749 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 840 840 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1267 1267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1064 1072 CLVTGKGH -> VRWEGCVFP (IN
FT VARSPLIC 1073 1629 MISSING (IN SHORT ISOFORM).
FT VARSPLIC 367 367 F -> L (IN REF. 1).
SQ SEQUENCE 1629 AA; 182649 MW; C1C4CEFF58B8941F CRC64;

Query Match 51.2%; Score 1046.5; DB 1; Length 1629;
Best Local Similarity 51.9%; Pred. No. 4.9e-79;
Matches 195; Conservative 57; Mismatches 109; Indels 15; Gaps 8;

QY 1 FVSIPRYVETLVVADESVMKPHGADLEHYLLTLTAARLYRHPSILNPINIVVKKVLL 60
DB 288 FLSYPRFVELVADNRMVSYHGENLOHYILTLMSIVASYIKDPSIGNLNINIVNLVI 347
QY 61 RDRDSGPKVTGNAALTLENFCAWQKLNKVSXKHPYWDYTAILFTRODLCA--TTCDTLG 119
DB 348 HNEQDPSISFNAQITLNFCAWQKLNKVSXKHPYWDYTAILFTRODLCA--TTCDTLG 404
QY 120 MADVTGMDPKRRCSVIEDDGLPSAFTAHGLHGFVNPHDKNVYKVEFGKLRANHMS 179
DB 405 LAELGTICDPYRSCSISDSGLSTAFTHAHLGFVNPHDKNVYKVEFGKLRANHMS 463
QY 180 PTLQIDRANPWSACSAIITDFELDSHGDCLLDQP--SKPISLPEDLPASYSILSOCEL 238
DB 464 PTLNFTYTPWMSKSRKYITFEFLDTGVGECLENEPSPYPLVPQLPGILYNVKNQCEL 523
QY 239 AFGVGSKPCPYMOYCTKLK---TGAKGOMVCQTRHFPWADGTSCEGKLCCLKACVER 295
DB 524 IFPGSGQCPYMQCRLWCVNNGVHKG---CRTQHTPWADGTECPGKHCKYGFQCPVK 580
QY 296 HNLNHRVDGSAKADPPGCSRTCCGGVGLARQCTNPTPANGKGYCEGVRYKRCNL 355
DB 581 -EMDVPVTDGWSGWSPTGCSRTCCGGGIGKTAIECNRNPEPKNGGKYCVGRMKRSCNT 639
QY 356 EPCPSASGKSPREQ 371
DB 640 EPCLKQK--RDRDREQ 653

RESULT 11
AT57_HUMAN STANDARD; PRT; 997 AA.
AC Q9UKP4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
GN with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.I., Hirohata S., Seidin M.F., Apte S.S.;
RT "ADAM-TS6, ADAM-TS5, and ADAM-TS7, Novel Members of a New Family of
RT Zinc Metalloproteases.";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -|- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -|- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,

```

```

CC -|- LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PFM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF140675; AAD56358.1; -.
CC HSP; P15167; IATL.
CC MEROPS; M12.231; -.
CC Genew; HGNC:223; ADAMTS7.
CC MIM; 605009; -.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_M12peptidse.
CC Pfam; PF00090; tsp_1; 1.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSP1; 2.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS50215; ADAM_MEPRO; 1.
CC PROSITE; PS50092; TSP1; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 BY SIMILARITY.
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match 40.18; Score 819; DB 1; Length 997;
Best Local Similarity 47.18; Pred. No. 2.2e-60;
Matches 177; Conservative 43; Mismatches 140; Indels 16; Gaps 10;

QY 2 VSIPIRYVETLVVADESVMKPHG--ADLEHYLLTLTAARLYRHPSILNPINIVVKKVLL 60
DB 238 VSKERKVELTVVADAKVVEYHGQVQVSVLTINMVAFLHDFPSIGNPHTITVRLVLL 297
QY 61 RDRDSGPKVTGNAALTLENFCAWQKLNKVSXKHPYWDYTAILFTRODLCA--TTCDTL 118
DB 298 EDEEEDLKITHADNTLTKSPCKWQKSNMKGDAPHLHDDTAILLTRKDLCAAMRNPCEL 357
QY 119 GMADVTGMDPKRRCSVIEDDGLPSAFTAHGLHGFVNPHD--NVKVCVEEFGKLRANHM 177
DB 358 GLSHVAGMCQPHRCSINEDTGLPLAFTVAHLGHSGFIQHDGSGNCEPV--GK--RPF 414
QY 178 MSPTLIQIDRANPWSACSAIITDFELDSHGDCLLDQPSKP--ISLPEDLPASYSILSOQC 236
DB 178 MSPTLIQIDRANPWSACSAIITDFELDSHGDCLLDQPSKP--ISLPEDLPASYSILSOQC 236

```

```
Db 415 MSPQLLYDAALPTWRSRCSQYITRFLDRGWGLCLDDPPAKDIIDFPSPVPGVLYDVSHQC 474
QY 237 ELAFGVGSKPCPYM-OYCTKLMCTCKAGQWVCOTRHPWADGTSCEGKLCCLKGACVER 295
Db 475 RLOYGAYSACFEDMDNVCHTWCSTGVT----TCHSKLDAADVGRGCGENKWLCSGECVP- 529
QY 296 HNLNKHVRDGNWAKWDPYGPCSRTCGGQVQLAROCNPTPTANGKYGCEGVYRVKYSRCL 355
Db 530 VGFPRPNDVGGHSGHSANSICSRSCMGVQSAEROCOTPTPKYKRGVCVGERKFRCLCNL 589
QY 356 EPCPSSASCKSPREQ 371
Db 590 QACP--AGRPSPRHVQ 603

RESULT 12
AT12_HUMAN STANDARD: PRT; 1593 AA.
ID AT12_HUMAN
AC P58397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;
RT "Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -1- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 8 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ250725; CAC20419.1;
CC Genew: HGNC:14605; ADAMTS12.
CC MIM: 606184;
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Reprolysin.
CC InterPro: IPR000884; TSPI.
```

```
DR InterPro: IPR000130; Zn_M12peptidse.
DR Pfam: PF00090; Tsp_1; 6.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR SMART: SM00209; TSPI; 8.
DR PROSITE: PS0215; ADAM_MEPRO; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE: PS50092; TSPI; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 240 BY SIMILARITY.
FT CHAIN 241 1593 ADAMTS-12.
FT DOMAIN 465 544 DISINTEGRIN-LIKE.
FT DOMAIN 545 596 TSP TYPE-1 1.
FT DOMAIN 597 700 CYS-RICH.
FT DOMAIN 701 826 SPACER 1.
FT DOMAIN 827 881 TSP TYPE-1 2.
FT DOMAIN 886 943 TSP TYPE-1 3.
FT DOMAIN 947 995 TSP TYPE-1 4.
FT DOMAIN 996 1315 SPACER 2.
FT DOMAIN 1316 1364 TSP TYPE-1 5.
FT DOMAIN 1367 1423 TSP TYPE-1 6.
FT DOMAIN 1426 1471 TSP TYPE-1 7.
FT DOMAIN 1426 1471 TSP TYPE-1 8.
FT DOMAIN 302 305 POLY-GLU.
FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 393 393 BY SIMILARITY.
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 105 105 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 951 951 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1104 1104 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1275 1275 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1300 1300 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1320 1320 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1503 1503 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 1593 AA; 177545 MW; 07F9F48E63BD83A3 CRC64;

Query Match 38.5%; Score 786; DB 1; Length 1593;
Best Local Similarity 43.1%; Pred. No. 2.1e-57;
Matches 163; Conservative 59; Mismatches 136; Indels 20; Gaps 10;

QY 2 VSIPRYVELLVVADDSMKVKEHAD-LEHYLLTLATAARLYRHPISILNPINIVVVKVLL 60
Db 242 ISKRWVETLVVADTKMIEYHGSVENSYILTIMNVTGLFHPNSIGNAHIVVRLILL 301
QY 61 RDRDSGPVKTGNAALTLNFCWAKKLNKVSXKHPYEDTAIFLTRODLCA--TTCDTL 118
Db 302 EEEQGLKIVHHAETLSLSSFCWKQKSNPKSDLPNVHVDVAVLLTRKDKICAGENRPCETL 361
QY 119 GNADVTGKCDPKRSCSVLEDGLPSAFTTAHELGHVFNMPHDNVK-VCESVFGKLRANHM 177
Db 362 GUSHLGMCQPHRSRCSINIEDSLPLAFTIAHELGHSGIQHDGKENDCEPVG---RHPI 418
QY 178 MSPTLIQIDRANPWSACSAAIITDFLOSGHGDCLLDQPSKP-ISLPEDLPFGASYTISQOC 236
Db 419 MSRQLQYDPTPLTWSKCSSEYITRFLDRGWGFCLLDIPKKKGLSKVIAFGVLYDVHHQC 478
QY 237 ELAFGVGSKPCPYM-OYCTKLMCTCKAGQWVCOTRHPWADGTSCEGKLCCLKGACVER 295
Db 479 QLOYGPNATFCOEVENVCOTLWCSSVKG----FCRSKLDAADAAAGTQCGEKKWCMAGKCI-- 532
QY 296 HNLNKHVRDGNWAKWDPYGPCSRTCGGQVQLAROCNPTPTANGKYGCEGVYRVKYSRSC 353
```

```

Db 533 -TVGKKPSIPGGMRGSPWHSRCSRTGAGVQSAERLCNNPEKFGKYCTGKRYRLC 591
QY 354 NLEPCPSSASGSKSFREQ 371
Db 592 NVHPCRSEA--PTFRQM 607

RESULT 13
AT10_HUMAN
ID AT10_HUMAN STANDARD; PRT; 1077 AA.
AC Q9H324;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA SEQUENCE FROM N.A.
RP Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT thrombospondin type I repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- COPACAPOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF163762; AAG35563.1; .
CC MEROPS; M12.235; .
CC Genew; HGNC:13201; ADAMTS10.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSPI.
CC InterPro; IPR000130; Zn_Mtpeptidse.
CC Pfam; PF00090; tsp.1; 5.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSPI; 5.
CC PROSITE; PS50215; ADAM_MERO; 1.
CC PROSITE; PS50092; TSPI; 2.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; zinc; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT NON_TER 1
FT PROPEP 1
FT CHAIN <1 207 BY SIMILARITY.
FT CHAIN 208 1077 ADAMTS-10.
FT METAL 366 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 367 366 ZINC (CATALYTIC).
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 CYS-RICH.

```

```

FT DOMAIN 680 802 SPACER.
FT DOMAIN 521 577 TSP TYPE-1 1.
FT DOMAIN 799 860 TSP TYPE-1 2.
FT DOMAIN 862 918 TSP TYPE-1 3.
FT DOMAIN 922 976 TSP TYPE-1 4.
FT DOMAIN 981 1031 TSP TYPE-1 5.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 769 769 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCBBF587 CRC64;

Query Match 36.8%; Score 751; DB 1; Length 1077;
Best Local Similarity 41.7%; Pred. No. 1.le-54;
Matches 163; Conservative 55; Mismatches 139; Indels 34; Gaps 11;

QY 2 VSIPIRYETLVVADESVMKPHG-ADLEHYLLTLLATAARLYRHPSILNPINIVVKKVLL 60
Db 209 VSRERYETLVVADKMMVAYHGRDRVEQYVLAIMNIVAKLFQDSGLSTVNIILVRLILL 268
QY 61 RDRDSGPKVTGNAALTNRNFCAMQKIL-----NKVSKRHPYWDITAILFTQDLC--G 111
Db 269 TEDQPTLEITHAGKSLDSFCKWQKSIIVNHSGHGNAPENGVAHDTAVLITRYDICIYK 328
QY 112 ATTCDTLMADVGWCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVKVCEVFG- 170
Db 329 NKPCGTLGLAFVGGMCERSCSYNEDIGLPOAFTIAHEIGTFTGMNHDGVCNSGARGQ 388
QY 171 ---KLRAHNMSPTLIOIDRANP--WSACAAIITFDLSDSHGDCLLDQPSK-PISLPE 224
Db 389 DPAKLMAAHITMKT-----NPFWSSCRDYITSLDGLGLCLNNRPRQDFVYPTV 441
QY 225 LPGASYTLSSQCELAFGVSGPKPYMOCYKWLCTGKAKGMVCQPRHFPWADGTCGEG 284
Db 442 APGOAYDAEQCRFQHGKYSQKQYGEVSELNCLSKSR---CITNSIPAAEGTLCQTH 498
QY 285 KL----CLKGACVERHNLNKHVRDGSNAKWDPYGCPSRTCTGGVQLARRCTNPTPANG 340
Db 499 TIDKGWCYKRYCVP-FGSRPEVDGANGPWTIPWGDSCRTCTGGGVSSSRHCDSPRTIGG 557
QY 341 KYCEGVRYKRSNLEPCPSSASGSKSFREQ 371
Db 558 KYCLGERRRHRSCTNDDCPGGS--QDFREQ 586

RESULT 14
AT52_HUMAN
ID AT52_HUMAN STANDARD; PRT; 1211 AA.
AC O95450;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
DE (Procollagen I/II amino-propeptide processing enzyme) (PCNP1)
DE (Procollagen I/II amino-propeptide processing enzyme) (PCNP1)
DE (Procollagen I/II amino-propeptide processing enzyme) (PCNP1)
DE (Procollagen I/II amino-propeptide processing enzyme) (PCNP1)
GN ADAMTS2 OR PCNP OR PCPNI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.
RP Tissue-Skin;
RX MEDLINE=99347935; PubMed=10417273;
RA Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,
RA Wertschke W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
RA Byers P.H., Lapierre C.M., Prockop D.J., Nussgens B.V.;
RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis
RT are caused by mutations in the procollagen I N-proteinase gene.";

```

Am. J. Hum. Genet. 65:308-317(1999).

-1- FUNCTION: Cleaves the propeptides of type I and II collagen prior to fibril assembly. Does not act on type III collagen. May also play a role in development that is independent of its role in collagen biosynthesis.

-1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains at Ala-1-Gln.

-1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

-1- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO COLLAGEN TYPE XIV (BY SIMILARITY).

-1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: LPNPI (SHOWN HERE) AND SPNPI; ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-PROCOLLAGEN PEPTIDASE ACTIVITY.

-1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.

-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

-1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).

-1- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos syndrome type VIIC (EDS-VIIC), a recessively inherited connective-tissue disorder characterized clinically by severe skin fragility and joint hypermobility and biochemically by the presence in skin of procollagen incompletely processed at the N-terminus.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

-1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

-1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.

-1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AJ003125; CAA05880.1; -

DR MEROPS: M12.301; -

DR Genew: HGNC:218; ADAMTS2.

DR MIM: 604539; -

DR MIM: 225410; -

DR InterPro: IPR001762; Disintegrin.

DR InterPro: IPR002870; Pep_M12B_propep.

DR InterPro: IPR001590; Reprolysin.

DR InterPro: IPR000884; TSP1.

DR InterPro: IPR000130; Zn_MTPeptdse.

DR Pfam: PF00090; tsp_1; 4.

DR Pfam: PF01421; Reprolysin; 1.

DR Pfam: PF01562; Pep_M12B_propep; 1.

DR SMART: SM00209; TSP1; 4.

DR PROSITE: PS00215; ADAM_MEPPO; 1.

DR PROSITE: PS50092; TSP1; 1.

DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.

DR PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; zymogen;

KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;

KW Alternative splicing; Ehlers-Danlos syndrome.

FT SIGNAL 1 29

FT PROPEP 1 29

FT CHAIN 30 253

FT METAL 254 1211

FT ACT_SITE 408 408

FT METAL 409 409

FT ACT_SITE 412 412

FT METAL 418 418

FT METAL 418 418

FT DOMAIN 480 560

FT DOMAIN 561 617

FT DOMAIN 618 722

FT SITE 691 693

FT SITE 723 851

FT DOMAIN 723 851

FT DOMAIN 852 911

FT DOMAIN 912 974

FT DOMAIN 975 1030

FT DOMAIN 40 43

FT DOMAIN 185 188

FT CARBOHYD 112 112

FT CARBOHYD 251 251

FT CARBOHYD 949 949

FT CARBOHYD 993 993

FT CARBOHYD 1031 1031

FT CARBOHYD 1098 1098

FT CARBOHYD 1145 1145

FT CARBOHYD 1150 1150

FT VARSPLIC 544 566

FT VARSPLIC 567 1211

FT SEQUENCE 1211 AA; 134722 MW; BECEEF25C3AD2D CRC64;

Query Match 31.0%; Score 633.5; DB 1; Length 1211;

Best Local Similarity 38.4%; Pred. No. 7.2e-45;

Matches 146; Conservative 55; Mismatches 142; Indels 37; Gaps 15;

QY 8 VETLVVADSMVKFHGAD-LEHVLTLTATAARLYRHPSILNPINIVVVKVLLLRDRSG 66

DB 268 IEVLGVDDSVVDFHGHKEHVQKYLTLMLNIVNEIYHDESLGAHINVLVRIILSYGKSM 327

QY 67 PKV-TGNAALTLRNCAWQKLNKVKSDKHPYVDYTAILETRDQLCGATTCDTLGMADVGT 125

DB 328 SLIEIGNPSQSLNVCWVAYLOQKPDTHGVHDHAIETLRQDF-GPSGMQ--GYAPVTG 384

QY 126 MCDPKKSCSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEVEFGKLRANHMMSPTLQ 184

DB 385 MCHPVSCTLNHEDGSSAFVVAHETGHLVMEHDCQGNRCGD---EVLGSIAPLVQA 441

QY 185 IDRANPWSACAAIITDFLDSDGHDCLDQP---SKPISLPDLPGASTLSQOCELARG 241

DB 442 AFHFHWSRCSQOELSRYLHS--YDCLDDPPAHDWP-ALPQ-LPLGHHYSMNEQCRDFEG 497

QY 242 VGSKPDPYMOY---CYKLWCTGKAKGMVQCTRHFPWADGTSCGEGKCLKGACV----- 293

DB 498 LGYMMCTAFRTDFPCQKLMCS-HPDNPFYCKTKKGKPLDGTMCAPGKHCFCIWLTPD 556

QY 294 --EHNLNKRVGDSWAKWDYPGVCSRTGGVQLARRQCTNTPANGGKYCGVVKYR 351

DB 557 ILKR-----DGSWGANSPFGSCSRTCTGTVKFRQCDNPHANGGRTCSGLAYDFQ 608

QY 352 SCNLEPCPSSASGSKSFREQ 371

DB 609 LCSHQDCPDSLA--DFREQ 626

RESULT 15

AT33_HUMAN STANDARD; PRT; 1205 AA.

ID AT33_HUMAN

AC O15072; Q9BXZ8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 3) (ADAM-TS 3) (ADAM-TS3) (Procollagen II amino-propeptide processing enzyme) (Procollagen II N-proteinase) (PC II-NP).

GN ADAMTS3 OR KIAA0366.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE OF 1-227 FROM N.A.

RX MEDLINE=21408482;

RA Fernandes R.J., Hirohata S., Engle J.M., Collige A., Cohn D.H.,

RA Eyre D.R., Apte S.S.;

FT *Procollagen II amino propeptide processing by ADAMTS-3. Insights on

dermatoparaxis";
 J. Biol. Chem. 276:31502-31509(2001).
 [2]
 SEQUENCE OF 5-1205 FROM N.A.
 TISSUE=Brain;
 MEDLINE=97349984; PubMed=9205841;
 Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 "Prediction of the coding sequences of unidentified human genes. VII.
 The complete sequences of 100 new cDNA clones from brain which can
 code for large proteins in vitro.";
 DNA Res. 4:141-150(1997).
 CC -!- FUNCTION: Cleaves the propeptides of type II collagen prior to
 fibril assembly. Does not act on types I and III collagens.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 matrix (by similarity).
 CC -!- TISSUE SPECIFICITY: Found in cartilage and skin.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AF247668; AAK28400.1; -;
 EMBL; AB002364; BAA20821.1; -;
 DR MEROPS; M12.220; -;
 Genew; HGNC:219; ADAMTS3.
 DR MIM; 605011; -;
 DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR002870; Pep_M12B_propep.
 DR InterPro; IPR001590; Repolysin.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR000130; Zn_MTPeptidse.
 DR Pfam; PF00090; tsp_1; 4.
 DR Pfam; PF01421; Repolysin; 1.
 DR Pfam; PF01562; Pep_M12B_propep; 1.
 DR SMART; SM00209; TSP1; 4.
 DR SMART; PS00215; ADAM_MEPRO; 1.
 DR PROSITE; PS50092; TSP1; 2.
 DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 KW Hydrolyase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 249 BY SIMILARITY.
 FT CHAIN 250 1205 ADAMTS-3.
 FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 399 399 BY SIMILARITY.
 FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 470 550 DISINTEGRIN-LIKE.
 FT DOMAIN 551 607 TSP TYPE-1 1.
 FT DOMAIN 608 712 CYS-RICH.
 FT DOMAIN 713 844 SPACER.
 FT DOMAIN 845 902 TSP TYPE-1 2.
 FT DOMAIN 903 965 TSP TYPE-1 3.
 FT DOMAIN 966 1017 TSP TYPE-1 4.
 FT DOMAIN 246 249 POLY-ARG.
 FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	345	345	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	475	475	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	814	814	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	942	942	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1205	AA; 135570	MW; EB07B286FC85FB87 CRC64;

Query Match 30.9%; Score 631.5; DB 1; Length 1205;
 Best Local Similarity 38.4%; Pred. No. 1.1e-44;
 Matches 144; Conservative 54; Mismatches 150; Indels 27; Gaps 13;

QY	8	VELTVADDSMYKFGHAD-LEHYLLTLLTAAARLYRHPHSILPINIVVVKVLLLRDRDSG	66
Db	258	IEVLGVDDSVVRFGKHEHVQNYLLTLMIVNEIYHDESIGVHINVLVRMIMLGAKSI	317
QY	67	PKV-TGNAALTIRNFCWQKLNKVS DKHPYWDTAIFLTRDLCGATTCDTLGMADYGT	125
Db	318	SLIERGNPSRSLNVCWRWASQQORSDLNHEHHDHAIFLTRDFGA---GMOGYPATVG	374
QY	126	MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHD---NVKYCEEVFGKLRANHHMSPTL	182
Db	375	MCHPVRSCITINHEDGFSFAFVVAHETGHLGMEHDGQGNRCGDETAMGS-----VMAPIV	429
QY	183	IQIDRANPKSACSAALITDLDGSHGDCILLDOP---SKPISLPEDLPGASYTLSSQCELA	239
Db	430	QAAFRHYHWSRCGGQELKRYIHS--YDCLLDDPFDDHDWP-KLPE-LPGINYSMEQCRFD	485
QY	240	FGVGSKPCPYMOY---CTKLWCTGKAKGMVQTRHFPWADGTSGEGKLCIKGACVERH	296
Db	486	FGVGKMCYTAFTDFDCKQLWCS-HPDNYPYCKTKKGLDGTCECAAGKWCYKGHCMWK-	543
QY	297	NLNKHRVDGSAKWDYPGCSRTCGGVOLARRQCTNPTTPANGKYCEGVVRVYKRSNLE	356
Db	544	NANQOKQDGNWGSWTGFGSCSRCTGTRFRTRQCNPNPMPINGGQDCPGVNFYQLCNTE	603
QY	357	PCPSSASGSKSREEQ	371
Db	604	ECQKHF--EDFRAQQ	616

Search completed: April 29, 2003, 17:14:00
 Job time : 9.56924 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:25 ; Search time 33.4317 Seconds
(without alignments)
2738.218 Million cell updates/sec

Title: US-10-009-332-1-copy_1_687

Perfect score: 3744

Sequence: 1 MILLLGLTLAFAGRTAGGFE.....RFDKCGVCGDKNCKKVTG 687

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3744	100.0	950	22	Human metalloprotease
2	3736	99.8	950	23	Human protease #2.
3	3725.5	99.5	928	23	Human metalloprotease
4	3725.5	99.5	952	23	Human metalloprotease
5	2497.5	66.7	505	21	Rat metalloprotease
6	1966	52.5	381	21	Human metalloprotease
7	1960.5	52.4	950	20	Human METH1 protei
8	1960.5	52.4	950	22	Human ADAM-type me
9	1960.5	52.4	950	22	Human METH1. Homo
10	1960.5	52.4	967	19	Human integrin lig

11	1960.5	52.4	968	22	AA550011	Protein; SEQ ID 12
12	1959.5	52.3	967	20	AA504142	Human Tango-71 pro
13	1947.5	52.0	896	21	AA212165	Mouse metalloprote
14	1938.5	51.8	967	20	AA78189	Human secreted pro
15	1921	51.3	950	21	AA53899	Amino acid sequenc
16	1913	51.1	367	23	AA22542	Human protease #3.
17	1751	46.8	727	20	AA78435	Human ADAMTS-1 pro
18	1699.5	45.4	890	20	AA49502	Human METH2. Homo
19	1699.5	45.4	890	22	AA50003	Human METH2. Homo
20	1696.5	45.3	889	22	AA74946	Human ADAM type me
21	1689.5	45.1	837	20	AA75425	Human aggrecan deg
22	1688.5	45.1	837	21	AA99429	Human PRO1563 (UNQ
23	1688.5	45.1	837	22	AAU2199	Human PRO polypept
24	1688.5	45.1	837	22	AA66178	Protein of the inv
25	1687.5	45.1	837	22	AA66178	Human aggrecanase-
26	1687.5	45.1	840	21	AA21256	Human metalloprote
27	1679	44.8	905	22	AA72284	Murine ADAMTS-8 am
28	1669	44.6	321	23	AAE22540	Human protease #1.
29	1611	43.0	930	22	AA72280	Murine ADAMTS-5 am
30	1588.5	42.4	870	21	AA21252	Rat metalloproteai
31	1587.5	42.4	930	20	AA75426	Human aggrecan deg
32	1580	42.2	929	21	AA841226	Human ORFX ORF990
33	1543.5	41.2	1629	23	ABG30703	Human aggrecanase
34	1543.5	41.2	1629	23	AAO14448	Human ADAMTS-SI pr
35	1543.5	41.2	1916	23	AAE19173	Human protease, PR
36	1543.5	41.2	1935	23	AAU72896	Human metalloprote
37	1540.5	41.1	1602	23	ABG30702	Human aggrecanase
38	1539.5	41.1	947	22	AA86950	Human metalloprote
39	1537.5	41.1	1073	21	AA821264	Human metalloprote
40	1523	40.7	1934	22	AA872301	Human ADAMTS-9 alt
41	1503.5	40.2	625	23	AA48394	Rat aggrecanase.
42	1472	39.3	1505	23	AAU72897	Human metalloprote
43	1466	39.2	680	21	AA21251	Human metalloprote
44	1460.5	39.0	1807	23	AAU71133	Human protease #12
45	1432	38.2	1882	22	AA872286	Human ADAMTS-9 ami

ALIGNMENTS

RESULT 1
AAGS2299
ID AAG622299 standard; protein; 950 AA.
XX AC AAG622299;
XX DT 23-AUG-2001 (first entry)
XX DE Human metalloprotease MDT56 protein.
XX KW Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
XX KW osteopathic; antiarthritic.
XX OS Homo sapiens.
XX PN WO200134785-A1.
XX PD 17-MAY-2001.
XX PF 10-NOV-2000; 2000WO-JP07917.
XX PR 11-NOV-1999; 99JP-0321740.
XX PR 16-MAY-2000; 2000JP-0144020.
XX FA (YAMA) YAMANOUCHI PHARM CO LTD.
XX FA (KAZU-) KAZUSA DNA RES INST.
XX PI Yamaaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX DR WPI: 2001-343602/36.
XX DR N-PSDB: AAH41003.
XX PT Metalloprotease with aggrecanase activity for treating joint diseases

```
PT especially osteoarthritis -
XX
PS Claim 1; Page 56-60; 85pp; Japanese.
XX
CC This invention relates to a metalloprotease with aggreganase activity.
CC The invention includes protein and DNA sequences of the metalloprotease,
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The present may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC the metalloprotease of the invention termed MDTs6.
XX
SQ Sequence 950 AA;

Query Match 100.0%; Score 3744; DB 22; Length 950;
Best Local Similarity 100.0%; Pred. No. 4.4e-285;
Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLGILTLAFAGRTAGGPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
DB 1 MLLGILTLAFAGRTAGGPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
QY 61 QEDFYHLTPDQAFAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAAVSLC 120
DB 61 QEDFYHLTPDQAFAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAAVSLC 120
QY 121 GGLRGAFGYRGAEEYISPLPNASAPAAQRNSOGAHLQRRGVPGSGDPTSRGCVASGW 180
DB 121 GGLRGAFGYRGAEEYISPLPNASAPAAQRNSOGAHLQRRGVPGSGDPTSRGCVASGW 180
QY 181 NPAILRALDPYKPRRAGFGESESRSSRAKRFVSIPIRYVETLVVADESMVKFHDLEH 240
DB 181 NPAILRALDPYKPRRAGFGESESRSSRAKRFVSIPIRYVETLVVADESMVKFHDLEH 240
QY 241 YLLTLATAARLYRHPSILNPINIVVVKLLLRDRSGPKVTGNAALTIRNFCAWQKKLN 300
DB 241 YLLTLATAARLYRHPSILNPINIVVVKLLLRDRSGPKVTGNAALTIRNFCAWQKKLN 300
QY 301 KVS DKHPEYWDTAILFTRODLGATTCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA 360
DB 301 KVS DKHPEYWDTAILFTRODLGATTCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA 360
QY 361 HELGHVFNMPHDNVKVEEVFGKLRANHMSPITLQIDRANPWSACSAIITDFLDSHG 420
DB 361 HELGHVFNMPHDNVKVEEVFGKLRANHMSPITLQIDRANPWSACSAIITDFLDSHG 420
QY 421 DCLLDQSPKIPISLPEDLPFGASYTLSSQCELAFCVGSKPCPYMOYCTKLWCTGKAKGMVC 480
DB 421 DCLLDQSPKIPISLPEDLPFGASYTLSSQCELAFCVGSKPCPYMOYCTKLWCTGKAKGMVC 480
QY 481 QTRHFPWADGTSCEGKCLIKGACVERHNLNKHVRVDSNAKMDPYGPCSRTCGGVQLAR 540
DB 481 QTRHFPWADGTSCEGKCLIKGACVERHNLNKHVRVDSNAKMDPYGPCSRTCGGVQLAR 540
QY 541 RQCTNPTPANGKYCEGVRVYKRSCLNLEPCPSASGKSPREOCEAFNGYNHSTNRLTA 600
DB 541 RQCTNPTPANGKYCEGVRVYKRSCLNLEPCPSASGKSPREOCEAFNGYNHSTNRLTA 600
QY 601 VAWPKYSGVSPRDKCKLICRANGTGYFYVLAPKVVVDGTLCSPDSTSVCVQGCIRAGCD 660
DB 601 VAWPKYSGVSPRDKCKLICRANGTGYFYVLAPKVVVDGTLCSPDSTSVCVQGCIRAGCD 660
QY 661 GNLSGKRRFDKCGVCGGDNKSKKVTG 687
DB 661 GNLSGKRRFDKCGVCGGDNKSKKVTG 687

RESULT 2
AAE22541
ID AAE22541 standard; Protein; 950 AA.
XX
AC AAE22541;
```

```
XX 26-JUL-2002 (first entry)
DT
XX Human protease #2.
DE
XX Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX
OS .Homo sapiens.
XX
FN WO200226949-A2.
XX
XX 04-APR-2002.
PD
XX
XX 27-SEP-2001; 2001WO-US30350.
PF
XX 29-SEP-2000; 2000US-236689P.
PR
XX (LEXI-) LEXICON GENETICS INC.
PA
PI Friddle CJ, Hilbun E;
XX
XX WPI; 2002-372123/40.
DR
XX N-PSDB; AAD35569.
DR
XX
PT Novel nucleic acid encoding a human protease, useful as a hybridization
PT probe for screening libraries and assessing gene expression patterns -
XX
XX Claim 6; Page 36-38; 41pp; English.
XX
CC The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridization probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 950 AA;

Query Match 99.8%; Score 3736; DB 23; Length 950;
Best Local Similarity 99.9%; Pred. No. 1.9e-284;
Matches 686; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLLGILTLAFAGRTAGGPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
DB 1 MLLGILTLAFAGRTAGGPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
QY 61 QEDFYHLTPDQAFAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAAVSLC 120
DB 61 QEDFYHLTPDQAFAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAAVSLC 120
QY 121 GGLRGAFGYRGAEEYISPLPNASAPAAQRNSOGAHLQRRGVPGSGDPTSRGCVASGW 180
DB 121 GGLRGAFGYRGAEEYISPLPNASAPAAQRNSOGAHLQRRGVPGSGDPTSRGCVASGW 180
QY 181 NPAILRALDPYKPRRAGFGESESRSSRAKRFVSIPIRYVETLVVADESMVKFHDLEH 240
DB 181 NPAILRALDPYKPRRAGFGESESRSSRAKRFVSIPIRYVETLVVADESMVKFHDLEH 240
QY 241 YLLTLATAARLYRHPSILNPINIVVVKLLLRDRSGPKVTGNAALTIRNFCAWQKKLN 300
DB 241 YLLTLATAARLYRHPSILNPINIVVVKLLLRDRSGPKVTGNAALTIRNFCAWQKKLN 300
QY 301 KVS DKHPEYWDTAILFTRODLGATTCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA 360
DB 301 KVS DKHPEYWDTAILFTRODLGATTCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA 360
QY 361 HELGHVFNMPHDNVKVEEVFGKLRANHMSPITLQIDRANPWSACSAIITDFLDSHG 420
DB 361 HELGHVFNMPHDNVKVEEVFGKLRANHMSPITLQIDRANPWSACSAIITDFLDSHG 420
```

Db 361 HELGHVFNPHDNVNVKCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 420
 QY 421 DCLLDQPSKPISLPEDLPASVTLTQQCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 480
 Db 421 DCLLDQPSKPISLPEDLPASVTLTQQCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 480
 QY 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHRRVDGSAKWDYPCSRCTCGGGVQLAR 540
 Db 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHRRVDGSAKWDYPCSRCTCGGGVQLAR 540
 QY 541 RQCTNPTPANGKYCEGVVRKYRSCNLEPCPSSASGKSFREDCQCAFNGYNHSTNRLTLA 600
 Db 541 RQCTNPTPANGKYCEGVVRKYRSCNLEPCPSSASGKSFREDCQCAFNGYNHSTNRLTLA 600
 QY 601 VAWVPKYSGVSPRDKCKLICRANGTYFYVLAPKVVDGTLGSPDSTSVCVGKCKIKAGCD 660
 Db 601 VAWVPKYSGVSPRDKCKLICRANGTYFYVLAPKVVDGTLGSPDSTSVCVGKCKIKAGCD 660
 QY 661 GNLGSKKRFDKCGVCGGDNKSKCKVTG 687
 Db 661 GNLGSKKRFDKCGVCGGDNKSKCKVTG 687
 RESULT 3
 AAU72899
 ID AAU72899 standard; Protein; 928 AA.
 AC AAU72899;
 DT 26-FEB-2002 (first entry)
 DE Human metalloprotease partial protein sequence #11.
 XX
 KW Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
 KW vasotropic; antihypertensive; analgesic; endocrine; neurotropic; tranquiliser;
 KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
 KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;
 KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;
 KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
 KW immune-related disease; cardiovascular disease; neuronal disease;
 KW migraine; sexual dysfunction; mood disorder; attention disorder;
 KW cognition disorder; hypotension; hypertension; psychotic disorder;
 KW dyskinesia; metabolic disorder; inflammatory disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200183782-A2.
 XX
 PD 08-NOV-2001.
 XX
 PF 04-MAY-2001; 2001WO-US14431.
 XX
 PR 04-MAY-2000; 2000US-201879p.
 XX
 PA (SUGEN-) SUGEN INC.
 PI Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 PI Payne V;
 XX
 DR WPI; 2002-041502/05.
 DR N-PSDB; AAS97182.
 XX
 PT Novel protease polypeptide useful for screening for substances that may
 PT be used to treat, e.g., cancers, immune-related diseases,
 PT cardiovascular disease, migraine, pain, psychotic and inflammatory
 PT disorders -
 XX
 PS Claim 28; Figure 2G; 232pp; English.
 XX
 CC The invention relates to an isolated, enriched, or purified protease
 CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
 CC screen for substances (S) that may modulate its activity. Administering
 CC S (which modulates protease activity in vitro) may be used to treat a

CC disease or disorder selected from cancers (e.g., of tissues, of blood or
 CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
 CC brain, ovarian, bladder or kidney), immune-related diseases and
 CC disorders, cardiovascular disease, brain or neuronal-associated diseases
 CC (e.g., central or peripheral nervous system diseases, migraine, pain,
 CC sexual dysfunction, mood disorders, attention disorders, cognition
 CC disorders, hypotension, hypertension, psychotic disorders, neurological
 CC disorders and dyskinesias), metabolic disorders and inflammatory
 CC disorders. (I) may also be useful as a diagnostic tool for a disease or
 CC disorder such as those above. AAU72876-AAU72910 represent human
 CC protease amino acid sequences of the invention.
 XX
 SQ Sequence 928 AA;
 Query Match 99.5%; Score 3725.5; DB 23; Length 928;
 Best Local Similarity 99.7%; Pred. No. 1.2e-283;
 Matches 686; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 1 MLLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPEDSGQGLIFQITAF 60
 Db 26 MLLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPEDSGQGLIFQITAF 85
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDFAAVSLC 120
 Db 86 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDFAAVSLC 145
 QY 121 GGLGAFGYRGAEVVISPLNAPAAQNSQGAHLQRRGVPGGSGDPTSRGCVASGW 180
 Db 146 GGLGAFGYRGAEVVISPLNAPAAQNSQGAHLQRRGVPGGSGDPTSRGCVASGW 205
 QY 181 NPAILRALDPYKPRRAGFESRRRRSGRAKRFVSPRYVETLVVADESVMKFGADLEH 240
 Db 206 NPAILRALDPYKPRRAGFESRRRRSGRAKRFVSPRYVETLVVADESVMKFGADLEH 265
 QY 241 YLTLTAARLYRHSILNPINIVVYVLLLRDRDGGPKVTGNAALTNRNFCAWQKLN 300
 Db 266 YLTLTAARLYRHSILNPINIVVYVLLLRDRDGGPKVTGNAALTNRNFCAWQKLN 325
 QY 301 KVSDDKHPEYWDTAILFTQDLCCATCTDLGMADVTGMDPKSCSVIEDDGLPSAFTTA 360
 Db 326 KVSDDKHPEYWDTAILFTQDLCCATCTDLGMADVTGMDPKSCSVIEDDGLPSAFTTA 385
 QY 361 HELGHVFNPHDNVNVKCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 420
 Db 386 HELGHVFNPHDNVNVKCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 445
 QY 421 DCLLDQPSKPISLPEDLPASVTLTQQCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 480
 Db 446 DCLLDQPSKPISLPEDLPASVTLTQQCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 505
 QY 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHRRVDGSAKWDYPCSRCTCGGGVQLAR 540
 Db 506 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHRRVDGSAKWDYPCSRCTCGGGVQLAR 565
 QY 541 RQCTNPTPANGKYCEGVVRKYRSCNLEPCPSSASGKSFREDCQCAFNGYNHSTNRLTLA 600
 Db 566 RQCTNPTPANGKYCEGVVRKYRSCNLEPCPSSASGKSFREDCQCAFNGYNHSTNRLTLA 625
 QY 601 VAWVPKYSGVSPRDKCKLICRANGTYFYVLAPKVVDGTLGSPDSTSVCVGKCKIKAGC 659
 Db 626 VAWVPKYSGVSPRDKCKLICRANGTYFYVLAPKVVDGTLGSPDSTSVCVGKCKIKAGC 685
 QY 660 GNLGSKKRFDKCGVCGGDNKSKCKVTG 687
 Db 686 GNLGSKKRFDKCGVCGGDNKSKCKVTG 713
 RESULT 4
 AAU74751
 ID AAU74751 standard; Protein; 952 AA.
 XX
 AC AAU74751;
 XX

DT 09-APR-2002 (first entry)

XX Human protease PRFS-11 protein sequence.

XX Human; protease; PRFS; gastrointestinal; Crohn's disease; cancer;

KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;

KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;

KW cell proliferative disorder; developmental disorder; epilepsy;

KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;

KW reproductive disorder; endometriosis.

XX Homo sapiens.

OS WO200198468-A2.

XX 27-DEC-2001.

XX 13-JUN-2001; 2001WO-US19178.

XX 16-JUN-2000; 2000US-212336P.

XX 22-JUN-2000; 2000US-213955P.

XX 29-JUN-2000; 2000US-215396P.

XX 07-JUL-2000; 2000US-216821P.

XX 14-JUL-2000; 2000US-218946P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;

PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;

PI Wallia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;

PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;

PI Kallack DA;

XX WPI: 2002-090437/12.

XX N-PSDB; ABK12894.

XX Twenty one human proteases (referred to as PRFS-1 to PRFS-21), useful

PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.

PT gastritis), cardiovascular (e.g. atherosclerosis) and cell

PT proliferative (e.g. cancer) disorders -

XX Claim 1; Page 144-146; 177pp; English.

XX The present invention relates to twenty one new human proteases,

CC referred to as PRFS-1 to PRFS-21. The PRFS polynucleotides and

CC polypeptides of the invention are useful in the diagnosis, treatment and

CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and

CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and

CC myocardial infarction, autoimmune/inflammatory e.g. acquired

CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell

CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker

CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.

CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and

CC endometriosis disorders. Numerous other examples of each disorder are

CC given in the specification. The present protein sequence represents

CC the human protease PRFS-11 protein of the invention.

XX Sequence 952 AA;

XX Query Match 99.5%; Score 3725.5; DB 23; Length 952;

XX Best Local Similarity 99.7%; Pred. No. 1.2e-283;

XX Matches 686; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MLLGILTLAFAGRTAGGPEREVVPIRLDPPDINGRRYWRGPDSDGGLIFQITAF 60

DB 1 MLLGILTLAFAGRTAGGPEREVVPIRLDPPDINGRRYWRGPDSDGGLIFQITAF 60

QY 61 QEDFYLHLTPDQAFAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDSFAAUSLC 120

DB 61 QEDFYLHLTPDQAFAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDSFAAUSLC 120

QY 121 GGLRGAFGYRGAEEYVISPLPNASAPAAQRNSQGAHLQRRGPGGSDPTSRGCVASGW 180

Db 121 GGLRGAFGYRGAEEYVISPLPNASAPAAQRNSQGAHLQRRGPGGSDPTSRGCVASGW 180

QY 181 NPATILRALDPYKPRRAGFSGESRRRRSGRAKRFRVSIIPRYVETLWVADESMVKFHGADLEH 240

Db 181 NPATILRALDPYKPRRAGFSGESRRRRSGRAKRFRVSIIPRYVETLWVADESMVKFHGADLEH 240

QY 241 YLLILLATAARLYRHPSTLNPINIVVVKVLLLRDRSDGPKVTGNAALTNRNFCAMQKKLN 300

Db 241 YLLILLATAARLYRHPSTLNPINIVVVKVLLLRDRSDGPKVTGNAALTNRNFCAMQKKLN 300

QY 301 KVSQKHPEYWDTAILFTRODLGCGATTCDTLGMADVGTMCDFKRSCSVIEDDGLPSAFTTA 360

Db 301 KVSQKHPEYWDTAILFTRODLGCGATTCDTLGMADVGTMCDFKRSCSVIEDDGLPSAFTTA 360

QY 361 HELGHVFNNPHDNNVKVCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDPLDSGHG 420

Db 361 HELGHVFNNPHDNNVKVCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDPLDSGHG 420

QY 421 DCLLDQPSKPISLPEDLPASYSYTLSSQCELAFAFGVSKPCPYMOCYTKLWCTGKAKGQWVC 480

Db 421 DCLLDQPSKPISLPEDLPASYSYTLSSQCELAFAFGVSKPCPYMOCYTKLWCTGKAKGQWVC 480

QY 481 QTRHFPWADGTSCGEGKLCGLKACVERHNLNKHVRVDSWAKWDPYGPCSRTCGGVQLAR 540

Db 481 QTRHFPWADGTSCGEGKLCGLKACVERHNLNKHVRVDSWAKWDPYGPCSRTCGGVQLAR 540

QY 541 RQCTNPTFANGKYCEGVRYRSCNLEPCPSSASGSKSFREEOCEAFNGYNHSTNRLTLA 600

Db 541 RQCTNPTFANGKYCEGVRYRSCNLEPCPSSASGSKSFREEOCEAFNGYNHSTNRLTLA 600

QY 601 VAMVPKYSVSPDRCKKLCIRANGTYFYVLAPK-VVDGTLCSPDSTSVCCVQGGKICAGC 659

Db 601 VAMVPKYSVSPDRCKKLCIRANGTYFYVLAPKVVVDGTLCSPDSTSVCCVQGGKICAGC 660

QY 660 DGNLGSKKRFRDKCGVCGGDNKSKKVTG 687

Db 661 DGNLGSKKRFRDKCGVCGGDNKSKKVTG 688

RESULT 5

AAAB21257

ID AAB21257 standard; Protein; 505 AA.

XX AAB21257;

XX 23-FEB-2001 (first entry)

XX Rat metalloproteinase ADAMTS-5.

XX Rat; ADAMTS-5; metalloproteinase; ADAM;

KW a disintegrin and metalloproteinase domain; thrombospondin domain;

KW vaccine; neutropic; neuroprotective; antiparkinsonian;

KW cerebroprotective; cyostatic; antiarthritic; immunosuppressive;

KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;

XX autoimmune disease; brain tumour; brain injury.

XX Rattus norvegicus.

XX WO2000053774-A2.

XX 14-SEP-2000.

XX 08-MAR-2000; 2000WO-US06237.

XX 08-MAR-1999; 99US-0264585.

XX (NEUR-) NEUROCRINE BIOSCIENCES INC.

XX kelner GS, Clark M, Maki RA;

XX WPI: 2000-594326/56.

XX N-PSDB; AAA95827.

PT Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX
PS Claim 12; Fig 14; 129pp; English.
XX
CC The present sequence is rat metalloproteinase ADAMTS-5. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMTS. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.
XX
SQ Sequence 505 AA;
Query Match 66.7%; Score 2497.5; DB 21; Length 505;
Best Local Similarity 93.5%; Pred. No. 1.4e-187;
Matches 462; Conservative 8; Mismatches 21; Indels 3; Gaps 2;
QY 89 QGLTGGSDLRRCFYSGDVNAEPDSFAVSLCGGLGAFYGAAYVISPLPNASAPAAQ 148
DB 13 QRLTSSLDLRRCFYSGVYNAEPDSFAVSLCGGLGAFYGAAYVISPLPNTSPEAQ 72
QY 149 RNSQAHLLQRRGVPGSGDPTSRGCVASGWNPAILLRALDPYKPRRAGFESRRRSRG 208
DB 73 RNSQAHLLQRRGVPGSGDPTSRGCVASGWNPAILLRALDPYKPRRTGVSCHNRRSG 132
QY 209 RAKRFSVPRVETLVVADESVMFVHGADLEHLLTLLATAARLYRHPSTLNINIVVYK 268
DB 133 RAKRFSVPRVETLVVADESVMFVHGADLEHLLTLLATAARLYRHPSTLNINIVVYK 192
QY 269 VLLLRDRSGPKVTGNAALTNRNFCAMOKLNVSKHPEYWDTAILFTQDLGCGATTC 328
DB 193 VLLLRDRSGPKVTGNAALTNRNFCAMOKLNVSKHPEYWDTAILFTQDLGCGATTC 252
QY 329 TLGMADVTCMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEFGKLRANH 388
DB 253 TLGMADVTCMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEFGKLRANH 312
QY 389 MMSPTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQPSKPTSLPDLPGASYTLSQQC 448
DB 313 MMSPTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQPSKPTSLPDLPGASYTLSQQC 372
QY 449 ELAFGVGSKPCPYMOCYTKLWCTGKAKGMVQCTRHFPWADGTSCGEGKLCGLGACVERH 508
DB 373 ELAFGVGSKPCPYMOCYTKLWCTGKAKGMVQCTRHFPWADGTSCGEGKLCGLGACVERH 432
QY 509 NLNKHVRVDSWAKNDPYGCSRTCGGGVQLARR--QCTNPTTPANGYKCYGVRVYKRSN 566
DB 433 NPNKYRVDPGWAKWEYPGCSRTCGGGAQLARRQVQATPLP-TGGYKCYGVRVYKRSN 491
QY 567 LEPCPSSASGSKFR 580
DB 492 LEPCPSSASGSKFR 505
RESULT 6
AAB21261
ID AAB21261 standard; Protein; 381 AA.
XX
AC AAB21261;
XX
DT 23-FEB-2001 (first entry)
XX
DE Human metalloproteinase ADAMTS-5.
XX
KW Human; ADAMTS-5; metalloproteinase; ADAM;

KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; nootropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 196 /note= "Xaa= any amino acid"
FT Misc-difference 268 /note= "Xaa= any amino acid"
FT Misc-difference 304 /note= "Xaa= any amino acid"
FT Misc-difference 308 /note= "Xaa= any amino acid"
FT Misc-difference 308 /note= "Xaa= any amino acid"
PN W0200053774-A2.
XX
PD 14-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US06237.
XX
PR 08-MAR-1999; 99US-0264585.
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX Kelner GS, Clark M, Maki RA;
XX WPI; 2000-594326/56.
DR N-PSDB; AAA95831.
XX
XX Polynucleotide encoding novel members of a disintegrin,
XX metalloproteinase and thrombospondin domain protein family used to
XX prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
PS Claim 12; Fig 23; 129pp; English.
XX
CC The present sequence is human metalloproteinase ADAMTS-5. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMTS. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.
XX
SQ Sequence 381 AA;
Query Match 52.5%; Score 1966; DB 21; Length 381;
Best Local Similarity 94.0%; Pred. No. 5.3e-146;
Matches 358; Conservative 3; Mismatches 18; Indels 2; Gaps 2;
QY 240 HYLLTLLATAARLYRHPSTLNINIVVYKLLLRDRSGPKVTGNAALTNRNFCAMOKKL 299
DB 2 HYRARAARAG-IFKHPSTLNINIVVYKLLLRDRSGPKVTGNAALTNRNFCAMOKKL 60
QY 300 NKVSDKHPEYWDTAILFTQDLGCGATTCDTLGMADVTCMDPKRSCSVIEDDGLPSAFTT 359
DB 61 NKVSDKHPEYWDTAILFTQDLGCGATTCDTLGMADVTCMDPKRSCSVIEDDGLPSAFTT 120
QY 360 AHELGHVFNMPHDNVKVCVEFGKLRANHMSPTLIQIDRANPWSACSAIITDFLDSGH 419
DB 121 AHELGHVFNMPHDNVKVCVEFGKLRANHMSPTLIQIDRANPWSACSAIITDFLDSGH 180
QY 420 GDCILLDQPSKPTSLPDLPGASYTLSQQCELAFGVGSKPCPYMOCYTKLWCTGKAKGMV 479
DB 181 GDCILLDQPSKPTSLPDLPGASYTLSQQCELAFGVGSKPCPYMOCYTKLWCTGKAKGMV 240

QY 480 CQTRHPWADGTCGEGKCLKACACVERHNLNKHVRDGSWAKWDPYGPCSRTCGGGVOLA 539
DB 241 CQTRHPWADGTCGEGKCLKACACVEXHNLNKHVRDGSWAKWDPYGPCSRTCGGGVOLA 300
QY 540 RQCTNPTP-ANGKYCEGVVKYRSCNLEPCPSASGKSPREOCEAFNGYNHSTNRLT 598
DB 301 RQXHOXPPLPGGYCEGVVKYRSCNLEPCPSASGKSPREOCEAFNGYNHSTNRLT 360
QY 599 LAVAWPKYSGVSPRDKCKLI 619
DB 361 LAVAWPKYSGVSPRDKCKLI 381
RESULT 7
AAY49501
ID AAY49501 standard; Protein; 950 AA.
XX
AC AAY49501;
XX
DT 10-JAN-2000 (first entry)
DE Human METH1 protein.
XX
KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
KW arterial-venous malformation; immune deficiency.
XX
OS Homo sapiens.
XX
PN WO9337660-A1.
XX
PD 29-JUL-1999.
XX
PF 22-JAN-1999; 99WO-US01313.
XX
PR 23-JAN-1998; 98US-0072298.
PR 28-AUG-1998; 98US-0098539.
XX
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
XX
PI IrueLA-Arispe L, Hastings GA, Ruben SM;
XX
XX WPI; 1999-590684/50.
DR N-PSDB; AA232000.
XX
PT New isolated metalloprotease thrombospondin polypeptides, useful for
PT treating hyperproliferative disorders, cancers or autoimmune disorders
XX
XX Claim 10: Fig 1; 457pp; English.
XX
XX AA232000 and AA232001 encode, and AAY49501 and AAY49502 represent, human
XX metalloprotease thrombospondin (METH) proteins METH1 and METH2
XX respectively. METH1 and METH2 have been found to be potent inhibitors of
XX angiogenesis both in vitro and in vivo. They can be used for treating
XX cancer and other disorders related to angiogenesis including abnormal
XX wound healing, inflammation, rheumatoid arthritis, psoriasis,
XX endometrial bleeding disorders, diabetic retinopathy, some forms of
XX macula degeneration, haemangiomas, and arterial-venous malformations.
XX They may be useful in treating deficiencies or disorders of the immune
XX system, by activating or inhibiting the proliferation, differentiation,
XX or mobilisation (chemotaxis) of immune cells. The etiology of these
XX immune deficiencies or disorders may be genetic, somatic, such as
XX cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
XX toxins), or infectious. They can also be used to treat inflammatory or
XX conditions, both chronic and acute conditions. The products can also be
XX used for detection and diagnosis. AA232002 to AA232080, and AAY49503 to

CC AAY49511 represent sequences given in the exemplification of the present
CC invention.
XX
SQ Sequence 950 AA;
Query Match 52.4%; Score 1960.5; DB 20; Length 950;
Best Local Similarity 53.6%; Pred. No. 4.7e-145;
Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;
QY 1 MLLLGILTLAFAGRTAGGFEPEVREVVPIRLDPDINGRRYRWGPEDSGQGLIFQITAF 60
DB 19 LLLLAALLAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTRLLRHAF 64
QY 61 QDEVYLHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRFCYSGDVNAEPDSF 114
DB 65 DQDLLELRPDSSFLAPGFTLQNGRKSGSETPLP-----ETDLAHCFYSGTVNGDPSSA 119
QY 115 AAVSLCGGLRGAFGYRGAEYVISPLPNAS-----APAAQRNSOGA-----HLIQ---RRGVPGG 165
DB 120 AALSCEGVGAFYLLGEAYFIQPLPAASERLATAAPGEKPPAPLPQPHLLRRNRQDVG 179
QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGFGESRRRRS 207
DB 180 TCGVVDDPRPTGKAETEDEGEDEGEQWMS-----PQDPALQGVGP-TGTGS 230
QY 208 GRAKFEVSIPIRYVEFLVVVADSMYKFGADLEHYLLTLATAARLYRHPILNPINIVV 267
DB 231 IRKRFVSHRYVETMLVADQSMAEFGSLKHYLLTFLSVAARLYKHPISIRNSVLVV 290
QY 268 KYLLLRDRSGPKVTGNAAALTNRNFCWOKKLNKVSQKHPYWDTAITLFRQDLCCATTC 327
DB 291 KILVHDEQKGEVTSNAALTNRNFCWQKHNPSPDRDAEHYDTALTFRQDLCSQTC 350
QY 328 DTLGMADVTMC DPRKCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEEFGKLKLAN 387
DB 351 DTLGMADVTMC DPRKCSVIEDDGLQAFAFTTAHELGHVFNMPHDDAKQCASLNGVQDS 410
QY 388 HWSPTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQSKPISLPELPGASYTLISQ 447
DB 411 HMASNLNLDHSQWSPCSAYMITSFLDNHGECLEMDKFNQPIQLPGDLPSTYDANRQ 470
QY 448 CELAFGVGSKPCP-YMOYCTKLWCTGKAKGQWCVOTRHFHWADGTCGEGKCLKACVE 506
DB 471 COFTGEDSKHCPDAASTCTLWCTGSGVLVQCTKHFPWADGTCGEGKWCINGKCVN 530
QY 507 RHNLNK--RVDGSAKWDYPGCSRTCGGVGVOLARROCTNPTTPANGGYCYGGRVRYR 563
DB 531 KYD-RKHFDTPFHGSGWGMGPGWDCSRTCGGVGVYTHRECDNPVPKNGGYCEGKRVYR 589
QY 564 SCNLEPCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRAN 623
DB 590 SCNLEDCPDN-NGKTFREOCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLICQAK 648
QY 624 GGYEYVLAPKVVDGTLCSPDSTSVCGKICKAGCNGLSKKRPFKDCVCGGDNKSK 683
DB 649 GIGYEFYVLPQKVVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKKFKDKCGVCGGNGSTCK 708
QY 684 KVTG 687
DB 709 KISG 712
RESULT 8
AAB73549
ID AAB73549 standard; Protein; 950 AA.
XX
AC AAB73549;
XX
DT 07-AUG-2001 (first entry)
XX
DE Human ADAM-type metalloprotease MDTs4, SEQ ID NO:4.
XX
KW Human; MDTs4; ADAM-type metalloprotease; drug screening;

KW A Disintegrin And Metalloprotease; cancer; arthritis.

OS Homo sapiens.

XX JP2001017183-A.

XX 23-JAN-2001.

XX 09-JUL-1999; 99JP-0196584.

XX 09-JUL-1999; 99JP-0196584.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

XX WPI; 2001-275950/29.

XX N-PSDB; AAB20224.

XX A new metal protease and its preparation for use as an anti-cancer and

XX anti-arthritis therapeutic -

XX Claim 1; Page 12-14; 22pp; Japanese.

XX The invention relates to the novel human ADAM (A Disintegrin And
CC Metalloprotease)-type metalloproteases MDTs4 (AAB73549) and MDTs5
CC (AAB73550). The metalloproteases can be used for the treatment of
CC cancers and arthritis. The invention also relates to the genes encoding
CC MDTs4 and MDTs5, vectors and host cells containing the MDTs4 or MDTs5
CC genes, the recombinant production of MDTs4 and MDTs5, and antibody
CC specific for MDTs4 or MDTs5, and methods of screening for compounds
CC which modulate the activity of MDTs4 and/or MDTs5. The present
CC sequence represents human MDTs4.

XX Sequence 950 AA;

Query Match 52.4%; Score 1960.5; DB 22; Length 950;
Best Local Similarity 53.6%; Pred. NO. 4.7e-145;
Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;

QY 1 MLLGLILTLAFAGRTAGGPEPREVVVPIRLDPDINGRIYWRGPDSDGDLFIQTAF 60
DB 19 LLLLAALLAVSDALGRPEDEELVVP-ELE-----RAP---GHGTTTLRLHAF 64
QY 61 QEDFYHLHLPDAQFLAPAFSTEHLG-----VPLQGLTGGSDLRRCFYSGDVNAEPDSF 114
DB 65 DQQLDLRLPDSFLAPGFTLQNVGKSGSETPLP-----ETDLAHCYFSGTVNGDPSSA 119
QY 115 AAVSLCGLRGAFYGAIEYVISPDPNAS---APAAQNSQGA---HLQ--RRGVPGG 165
DB 120 AALSCEGVGRGAFYLLGEAYFIQPLPAASERLATAAPGKPPAPLQFHLRRNRQGDVGG 179
QY 166 PSG-----DPTSRC-----CVASG--WNPAILRALDPYKPRAGFSGESRRRS 207
DB 180 TCGVDDDEPPTKATEDEDEGEDGEPQWS-----PQDPALQGVGP--TGTS 230
QY 208 GRAKRFVPIRYVETLVVADESVMKFGADLEHLYLITLATAARLYRHPSPILNINIVV 267
DB 231 IRKRFVSHRYVETMLVADQSMAREHSGSLAHYLLTFSVAARLYKHPISIRNSVSLVV 290
QY 268 KVLRLDRDSGPVKTGNAALTNLNFCWAKKLNKVDKHPYWDYTAFLTRODLCGATTC 327
DB 291 KILVIHQEQGPVTSNAALTNLNFCWQKQHPNPPSDRDAEHYDTAFLTRODLCGATTC 350
QY 328 DTLCMDVGTMDPKKSCSVIEDGLPSAFTTAHELGHVFNPHNPNHVNCEVFGKLRAN 387
DB 351 DTLCMDVGTMDPKKSCSVIEDGLPSAFTTAHELGHVFNPHNPHNPHDQKQASLNGVND 410
QY 388 HMTSPTLIQIDRANPNSACSAIITDFLDSGHGDCLLDQPSKPISLPEDLPASVTLT 447
DB 411 HMTASMLNLDHQPNSPCSAIYITSLDNGHGECLMDKPNQPIQLPGLDPLCTSDANRQ 470
QY 448 CELAFGVGSKPCP-YMQYCTKWKCTGKAGQWVCOTRHPFWADGTCGEGKLCAGAC 506
DB 471 CQFTFGEDSKHCPDAAASTLTCTGTSGVLVCQTKHPFWADGTCGEGKWCINGKCVN 530

QY 507 RHNLAKH---RVDGSWAKWDYPGCSRTCCGGVQLARRQCTNPTPANGKYCEGVRYK 563
DB 531 KTD-RKHFDTPPHGSGWGMGPMGDCSRTCCGGVQVTMRCDNPVKNGKYCEGKRYR 589
QY 564 SCNLEPCPSSASCKSFRECEAFNGYNHSTNRLTLAVAWVPKYSGVSPRDKCLICRAN 623
DB 590 SCNLEDCPDN-NGKTFREECEAHNEFFSKASFGSGPAAVEIPKYAGVSPKDRCKLICQAK 648
QY 624 GTGYFVVLAPKVVVDGTLCSPDSTSVYQVKCIKAGCDGNLGSKRFRDKCGVCGDGNK 683
DB 649 GIGYFVLQPKVVVDGTPCSPDSTSVYQVQCVKAGCDRIIDSKKKFKDKCGVCGGNGSTCK 708
QY 684 KVTG 687
DB 709 KISG 712

RESULT 9

AAB50002

ID AAB50002 standard; Protein; 950 AA.

XX AC AAB50002;

XX DT 19-MAR-2001 (first entry)

XX DE Human METH1.

XX Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma; trachoma; vascular adhesion; myocardial angiodysplasia;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiofibroma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control.

XX Homo sapiens.

OS WO200071577-A1.

PN 30-NOV-2000.

XX 25-MAY-2000; 2000WO-US14462.

XX 25-MAY-1999; 99US-0318208.

PR 20-JUL-1999; 99US-0144882.

PR 10-AUG-1999; 99US-0147823.

PR 13-AUG-1999; 99US-0373658.

PR 22-DEC-1999; 99US-0171503.

PR 22-FEB-2000; 2000US-0183792.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA (IRUE/) IRUELA-ARISPE L.

PA (HAST/) HASTINGS G A.

PA (RUBE/) RUBEN S M.

PA (JONA/) JONAK Z L.

PA (TRUL/) TRULLI S H.

PA (FORN/) FORNWALD J A.

PA (TERE/) TERRETT J A.

XX IrueLa-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Fornwald JA, Terrett JA;
XX WPI; 2001-025136/03.
DR N-PSDB; AAC90057.
XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT

PT rheumatoid arthritis and psoriasis -

XX Claim 15; Fig 1; 768pp; English.

XX The present sequence is human METH1 (ME for metalloprotease and TH for thrombospondin). METH1 can be used for inhibiting angiogenesis in an individual, and for treating cancer, benign tumours, an ocular angiogenic disease, rheumatoid arthritis, psoriasis, delayed wound healing, nonunion endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion fractures, scleroderma, trachoma, vascular adhesions, myocardial angiogenesis, coronary collateral, cerebral collaterals, arteriovenous malformation, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque neovascularisation, telangiectasia, haemophilic joints, angiofibroma, fibrovascular dysplasia, wound granulation, Crohn's disease or atherosclerosis. METH1 can also be used in birth control. METH1 can also be used in diagnostic methods for the prognosis of cancer.

XX Sequence 950 AA;

Query Match 52.4%; Score 1960.5; DB 22; Length 950;
Best Local Similarity 53.6%; Pred. No. 4.7e-145;
Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;

QY 1 MLLGILTLAFAGTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
DB 19 LLLAALLAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTRRLRHAF 64
QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGSSDLRRCFYSGDYNAPDSF 114
DB 65 DQQLDLRLPDSFLAPGFTLQNGRKSGSETPLP-----ETDLAHCFTSGTVNGDPSSA 119
QY 115 AAVSLCGGLRGAFVGRGAEYVISPLPNAS---APAAORNSOGA-----HLLQ---RRGVPG 165
DB 120 AALSCEGVGRGAFYLLGAYFIQPLPAASERLATAAAGEKPPAPLOPHLLRRNRQDVG 179
QY 166 PSG-----DPTSRC-----GVASG---WNPAIRLALDPYKPRRAGFESRRRS 207
DB 180 TCGVVDDPRPTGKAETEDEGTEDEGPQWS-----PODPALQGVGP-TGTGS 230
QY 208 GRAKFSVSTPRVETLVVADESQWKFHCADLEHYLLTLATAARLYRHSILNPIV 267
DB 231 IRKKFVSHRYVETMLVADQSMAEFFHSGGLKHYLLTLFVAAKLYKHSIRNSVLVV 290
QY 268 KYLLLRDRDSDGPKVTGNAAFLRNFCAWQKKLNKVSOKHPEYWDATILFTRQDLGCATTC 327
DB 291 KILVTHDEQKPEVTSNAALILRNFCAWQKQHNPPSDRDAEHVDTALFTRQDLGSGQTC 350
QY 328 DTGLMADVTGCDPKRSCVIEDDGLPSAFTAHGLGHVFNHNDVKNVCEEVFGKLRAN 387
DB 351 DTGLMADVTGCDPKRSCVIEDDGLPSAFTAHGLGHVFNHNDVKNVCEEVFGKLRAN 410
QY 388 HMASPTLIQIDRANPWSACSAIITDFLDGSGDCLLDQPSKPTSLPDLPGASYTL 447
DB 411 HMASMLNLDHSDQWPCSAAYMITSFLDNGHCELMKQNPQLPGLPGTSDANRQ 470
QY 448 CELAFGVSGKPCP-YMQVCTKLWCTGRAGQMVQOTRHFPAWQTSCEGKCLKACV 506
DB 471 COFTFEGDSKPCDAASTCTLMCTGTSGGVVLCQTHFPAWQTSCEGKCLKACV 530
QY 507 RNLANKH---RVDSGWAKWDVPGPCRTCCGGGVOLARROCTNPTPANGKVCYGRV 563
DB 531 KTD-RKHEDTFPHGSGWGWGFWGDCSRTCCGGGVQYTHRECDNVPKNGGKVCGRV 589
QY 564 SCNLEPCPSSAGSKSFREOCEAFNGYNHSTNRLTLAVAPVKYSVSPRDKCKLCIRAN 623
DB 590 SCNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSGPAVEMIPKYAGVSPKDRCKLCQAK 648
QY 624 GTGYEYVLAPKVVDTGLCSPTSVCVQKCIKAGCDNGLSKRRFDKCGVCGDNKSK 683
DB 649 GIGYEFVLQPKVVDGTPCSPDSTSVQVQCGVCRAGCDRIIDSKKKFDKCGVCGDNKSK 708
QY 684 KVTG 687
|::|

DB 709 KISG 712

RESULT 10
AAW80285
ID AAW80285 standard; Protein; 967 AA.
AC AAW80285;
XX 19-JAN-1999 (first entry)
XX Human integrin ligand polypeptide ITGL-TSP.
XX ITGL-TSP; integrin ligand; angiogenic disease; cancer; atherosclerosis;
XX chronic inflammatory disorder; rheumatoid arthritis; tissue remodeling;
XX macular degeneration; diabetic retinopathy; Alzheimer's disease; human;
XX restenosis.
XX Homo sapiens.
XX EP874050-A2.
XX 28-OCT-1998.
XX 27-JAN-1998; 98EP-0300575.
XX 24-APR-1997; 97US-0845496.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.
XX Fronwald JA, Hastings GA, Jonak ZL, Terrett JA;
PI Trull SH;
XX WPI; 1998-544643/47.
XX N-PSDB; AAV66508.
XX DNA encoding integrin ligand polypeptide ITGL-TSP - used to treat
XX angiogenic diseases, restenosis, Alzheimer's disease and in tissue
XX remodeling
XX Claim 11; Pages 6-9; 24pp; English.
XX This represents a human integrin ligand polypeptide ITGL-TSP. ITGL-TSP
XX polypeptides can be used in the treatment of angiogenic diseases such as
XX cancer, cancer metastasis, chronic inflammatory disorders, rheumatoid
XX arthritis, atherosclerosis, macular degeneration or diabetic retinopathy,
XX restenosis, Alzheimer's disease and tissue remodeling. They can be used
XX to treat a subject in need of enhanced activity or expression of the
XX ITGL-TSP polypeptide.
XX Sequence 967 AA;

Query Match 52.4%; Score 1960.5; DB 19; Length 967;
Best Local Similarity 53.6%; Pred. No. 4.8e-145;
Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;

QY 1 MLLGILTLAFAGTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
DB 36 LLLAALLAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTRRLRHAF 81
QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGSSDLRRCFYSGDYNAPDSF 114
DB 82 DQQLDLRLPDSFLAPGFTLQNGRKSGSETPLP-----ETDLAHCFTSGTVNGDPSSA 136
QY 115 AAVSLCGGLRGAFVGRGAEYVISPLPNAS---APAAORNSOGA-----HLLQ---RRGVPG 165
DB 137 AALSCEGVGRGAFYLLGAYFIQPLPAASERLATAAAGEKPPAPLOPHLLRRNRQDVG 196
QY 166 PSG-----DPTSRC-----GVASG---WNPAIRLALDPYKPRRAGFESRRRS 207
DB 197 TCGVVDDPRPTGKAETEDEGTEDEGPQWS-----PODPALQGVGP-TGTGS 247

QY 208 GRAKRFVSIPIRYETLTVVADSMVKHGADEHYLLTLLATAARLYRHPSILNPIVIVV 267
Db 248 IRKRFVSSHRYETMLVADQSMKAEHSGGLKHYLLTSLVAARLYRHPSIRNSVSLVVV 307
QY 268 KVLRLDRDSDGPKVTGNAAITLRFNCAWKKLNKSDKHPEYWDYTAILETRDLCGATTC 327
Db 308 KILVIHDEQKGEPEVTSNAALTFLNFCNWKQKHNPPSDRDAEHYDTAILFTRODLGSGTQC 367
QY 328 DTGLMADVTWCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGKLRAN 387
Db 368 DTGLMADVTWCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGKLRAN 427
QY 388 HNSPTLIQIDRANPWSACSAIITDFLDGSHGDCILLDOPSKPISLPEDLPASGYTLSSQ 447
Db 428 HMASMLNLDHDSOPSPSCSAYMITSLDNGHGECLMDKPNQPIQLPGDLPISYDANKQ 487
QY 448 CELAFGVGSKPCP-YMOYCTKLWCTGKAKQMVCOYTHFFPWADGTCGSKLCKGACVE 506
Db 488 CQFTFGEDSKHCPDAASTCTLWCTGTSGVLVCQTKHFPWADGTCGSGKWCINGKVCN 547
QY 507 RHNLNKH---RVDSWAKWDPYPCSTCGGQVQLARRQCTNTPANGKCYCBGVRYKVR 563
Db 548 KTD-RKHFDPFHFGSGWGMGFWDCSRTCGGVQYTMRECDNVPKNGKCYCBGVRYKVR 606
QY 564 SCNLEPCPSASGSKFRECEAFNGYNHSTNLTILAVAWPKYSGVSPRDKCKLICRAN 623
Db 607 SCNLEDCPDN-NGKTFREECEAHNEFSKASFGSGFAVEWPKYAGVSPADKCLICQAK 665
QY 624 GTGYFVYVLPKVVVDGTLCPDSTSVCVQGVKICIRAGCDGNLGSKKRFDKGVCGDGNKSK 683
Db 666 GIGYFVFLQPKVVVDGTPCSPDSTSVCVQGVKICIRAGCDRIIDSKKFDKGVCGDGNSTCK 725
QY 684 KVTG 687
Db 726 KISG 729

RESULT 11
AAB50011
ID AAB50011 standard; Protein; 968 AA.
XX AC AAB50011;
XX DT 19-MAR-2001 (first entry)
XX DE Protein; SEQ ID 125.
XX KW Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiofibroma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control.
XX OS Homo sapiens.
XX PN W0200071577-A1.
XX PD 30-NOV-2000.
XX PF 25-MAY-2000; 2000WO-US14462.
XX PR 25-MAY-1999; 99US-0318208.
XX PR 20-JUL-1999; 99US-0144882.
XX PR 10-AUG-1999; 99US-0147823.
XX PR 13-AUG-1999; 99US-0373658.
XX PR 22-DEC-1999; 99US-0171503.
XX PR 22-FEB-2000; 2000US-0183792.

XX PA (HUMA-) HUMAN GENOME SCI INC.
PA (SMK) SMITHKLINE BEECHAM CORP.
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUB/) RUBEN S M.
PA (JONA/) JONAK Z L.
PA (TRUL/) TRULLI S H.
PA (FORN/) FORNWALD J A.
PA (TERR/) TERRETT J A.
XX PI IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Fornwald JA, Terrett JA;
XX DR WPI; 2001-025136/03.
XX PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX PS Claim 15; Pages 759-763; 768pp; English.
XX CC The present invention relates to human METH1 and METH2 (ME for
CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC METH can be used for inhibiting angiogenesis in an individual, and for
CC treating cancer, benign tumours, an ocular angiogenic disease,
CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH can also be used in birth control. METH can also
CC be used in diagnostic methods for the prognosis of cancer. The present
CC sequence is a protein isolated in the present invention.
XX SQ Sequence 968 AA;
Query Match 52.4%; Score 1960.5; DB 22; Length 968;
Best Local Similarity 53.6%; Pred. No. 4.8e-145;
Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;
QY 1 MLLGILTLAPAGTAGGFEPEVVPVIRLDPDINGRYYWRGPEDSGDLGFIQTAF 60
Db 37 LLLAAALLAVSDALGRPSEEDLVVP-ELE-----RAP---GHGTTTLRLHAF 82
QY 61 QEDFYHLHTPDQAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFCYSGDVNAEPDSF 114
Db 83 DQQLDLELRPDSSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFTGTVNGDPSSA 137
QY 115 AAVSLCGGLRGAFGRGAEYVISPLPNAS---APAAQNSOGA-----HLLQ--RGVPGG 165
Db 138 AALSCEGVGRGAFYLLGEAYFIQPLPAASERLATAAPGKPPAPLQFHLRLNRQGDVG 197
QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPVKPRAGGSESRSSRS 207
Db 198 TCGVVDDEPRFTGKAETEDEGEDEGEQWSS-----PODPAQGVGP-TGTGS 248
QY 208 GRAKRFVSIPIRYETLTVVADSMVKHGADEHYLLTLLATAARLYRHPSILNPIVIVV 267
Db 249 IRKRFVSSHRYETMLVADQSMKAEHSGGLKHYLLTSLVAARLYRHPSIRNSVSLVVV 308
QY 268 KVLRLDRDSDGPKVTGNAAITLRFNCAWKKLNKSDKHPEYWDYTAILETRDLCGATTC 327
Db 309 KILVIHDEQKGEPEVTSNAALTFLNFCNWKQKHNPPSDRDAEHYDTAILFTRODLGSGTQC 368
QY 328 DTGLMADVTWCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGKLRAN 387
Db 369 DTGLMADVTWCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGKLRAN 428
QY 388 HNSPTLIQIDRANPWSACSAIITDFLDGSHGDCILLDOPSKPISLPEDLPASGYTLSSQ 447

[illegible]

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 11.1439 Seconds
(without alignments)
1813.869 Million cell updates/sec

Title: US-10-009-332-1-copy_1_687

Perfect score: 3744
Sequence: 1 MLLGLTILAFAGTAGGFE.....RPDKCGVCGDNKSKCKVTG 687

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1960.5	52.4	967	4	US-09-130-491-2
2	1689.5	45.1	837	4	US-09-122-126B-2
3	1679	44.8	905	4	US-09-369-364A-9
4	1611	43.0	930	4	US-09-369-364A-2
5	1587.5	42.4	930	4	US-09-122-126B-15
6	1536	41.0	608	4	US-09-130-491-13
7	1432	38.2	1882	4	US-09-369-364A-13
8	1394.5	37.2	874	4	US-09-369-364A-15
9	1077	28.8	997	4	US-09-369-364A-7
10	1072	28.6	551	4	US-09-130-491-16
11	1004	26.8	518	4	US-09-369-364A-22
12	976	26.1	1211	4	US-09-491-522-5
13	968.5	25.9	1224	4	US-09-930-872-4
14	944.5	25.2	1205	4	US-09-491-522-11
15	930.5	24.9	1081	4	US-09-369-364A-17
16	905.5	24.2	859	4	US-09-369-364A-5
17	841	22.5	481	4	US-09-130-491-8
18	662.5	17.7	245	4	US-09-369-364A-11
19	576.5	15.4	566	4	US-09-491-522-7
20	441	11.8	491	4	US-09-930-872-2
21	358	9.6	525	4	US-09-369-364A-21
22	346.5	9.3	812	4	US-09-632-098-4
23	339	9.1	802	4	US-09-632-098-2
24	303	8.1	464	4	US-09-411-329C-14
25	301.5	8.1	616	4	US-09-608-730-1
26	301	8.0	462	4	US-09-411-329C-3
27	301	8.0	462	4	US-09-411-329C-17

28	301	8.0	529	2	US-08-836-442-3	Sequence 3, Appli
29	285	7.6	621	4	US-09-026-001A-6	Sequence 6, Appli
30	278	7.4	621	4	US-09-026-001A-18	Sequence 18, Appli
31	277.5	7.4	814	4	US-09-813-819-4	Sequence 4, Appli
32	277.5	7.4	814	4	US-09-920-048-4	Sequence 4, Appli
33	277.5	7.4	855	4	US-09-813-819-2	Sequence 2, Appli
34	277.5	7.4	855	4	US-09-920-048-2	Sequence 2, Appli
35	273	7.3	613	4	US-09-026-001A-10	Sequence 10, Appli
36	269.5	7.2	769	1	US-08-243-542-4	Sequence 4, Appli
37	269.5	7.2	769	1	US-08-477-407-4	Sequence 4, Appli
38	269.5	7.2	769	1	US-08-484-355-4	Sequence 4, Appli
39	259	6.9	592	4	US-09-026-001A-14	Sequence 14, Appli
40	253.5	6.8	670	1	US-08-243-542-3	Sequence 3, Appli
41	253.5	6.8	670	1	US-08-477-407-3	Sequence 3, Appli
42	253.5	6.8	670	1	US-08-484-355-3	Sequence 3, Appli
43	248.5	6.6	1170	1	US-08-313-288B-20	Sequence 20, Appli
44	244	6.5	335	4	US-09-152-060-64	Sequence 64, Appli
45	241.5	6.5	751	2	US-08-836-443-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130.491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-130-491-2

Query Match	52.4%	Score 1960.5;	DB 4;	Length 967;
Best Local Similarity	53.6%	Pred. No. 9.6e-156;		
Matches	388;	Conservative	91;	Mismatches 178;
				Indels 67;
				Gaps 17;
QY	1	MLLGLTILAFAGTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAF	60	
DB	36	LILLAAALAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTTTLRLHAF	81	
QY	61	QEDFYHLTPDQAFAPAFSTEHLG-----VPLQGLTGGSSDLRRRCFYSGDVNAEPDSF	114	
DB	82	DQDLDELRPDSSFLAPGETLQNVGRKSGSETPLP-----ETDLAKCEYSTVNGDPSA	136	
QY	115	AAVSLGGGLRGAFGRGAEYVISPFPNAS---APAAQNSOGA---HLLQ--RRGVPG	165	
DB	137	AALSCEGVGRGAFYLLGEAYFTQPLPAASERLATAAPGKPPAPLQFHLLRRNRQDVG	196	
QY	166	PSQ-----DPTSRC-----GVASG--WNPAIRLALDPYKPRAGFSGSRRS	207	
DB	197	TCGVVDEPRPTGKAETDEDEGTEGEGPOWS-----PQDPALQGVGP-TGTGS	247	
QY	208	GRAKREVSIPRYVETLWVADESMVKFHGADLEHYLLTLLATAARLYRHPSTLNPINIVV	267	
DB	248	IRKKREVSRRYVETMLVADQSMAEFHGSLKHYLLTLFVAAARLYKHPSTRNSVSLVV	307	
QY	268	KVLLLRDROSGPKYTCNAALTIRNFCANWKKLNKYSDKHPEYWDPAIFLTQDLCGATTC	327	
DB	308	KILVTHDEQGPVTSNAALTIRNFCNWKQHNPPSDRDAEHYDPAIFLTQDLCGSGTC	367	

QY 328 DTLMADVGTWCPKRCSCSVIEDDGLPSAFTTAHELGHVFNPHNDNPKYCEVFGKLRAN 387
Db 368 DTLMADVGTWCPKRCSCSVIEDDGLQAAFTTAHELGHVFNPHNDNPKYCEVFGKLRAN 427
QY 388 HNSPTLIQIDRANPWSACSAIIIDFDSHGDCLLDQPSKPSLPLDPLGASVYLSQ 447
Db 428 HNSMNLNDSQWSPSCSAIYIISFLDNGHGECLMDKPNQIPLDPLGASVYLSQ 487
QY 448 CELAFGVGSKPCP-YMYCTKLMCTGKAGKQWVCOTRHPWADGTSCGEGKCLKLGACVE 506
Db 488 CQTFEGEDSKHCPDAASTCTWCTGTSGVLVQCOTKHPWADGTSCGEGKCLKLGACVE 547
QY 507 RHNLNKHH--RVDSWAKMDPSPGSCRTSCGGVQLARQCTNPTPANGKCYCEGVYKYR 563
Db 548 KTD-RKHEDTPHSGWGMWPGWDSRTSCGGVQVYTMEDCNPPVKNKGKCYCEGVYKYR 606
QY 564 SCHLEPCPSASGSKSPREOCQAFNGYNHSTNRLTLAVAWPKYSGVSPDRCKLICRAN 623
Db 607 SCHLEPCPN-NGKTFREOCQAFNEFSKASFGSGPAVEWIPKAVGSPDRCKLICQAK 665
QY 624 GTGYFYVLAPKVVDTLCSPDSTSCVQKCIKAGCDNLGSKRFDKCGVCGDGNKSK 683
Db 666 GIGYFYVLAPKVVDTLCSPDSTSCVQKCIKAGCDNLGSKRFDKCGVCGDGNKSK 725
QY 684 KVTG 687
Db 726 KISG 729

RESULT 2

US-09-122-1268-2

; Sequence 2, Application US/091221268

; Patent No. 6451575

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES

; FILE REFERENCE: DM6909

; CURRENT APPLICATION NUMBER: US/09/122.1268

; CURRENT FILING DATE: 1998-07-24

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 837

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-122-1268-2

Query Match 45.1%; Score 1689.5; DB 4; Length 837;
Best Local Similarity 47.2%; Pred. No. 4e-133;
Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;

QY 1 MLLGLTLTAFAGTAGGFEPEREVVPRLDPDINGRRYWRGPDSDGGLIFQITAF 60
Db 37 LLLLLLTLTAFAGTAGGFEPEREVVPRLDPDINGRRYWRGPDSDGGLIFQITAF 88
QY 61 QEDFYVLPDQAFAPAFSTHGLVPLQGLTGGSSDLRRRCFYSGDVNAEPDFAAVSLC 120
Db 89 GETLLELQDQSGVQVGLTVQLGAPE-LLGGAEP--GTYLGTTINGDPESVSLHWD 145
QY 121 GG-LRGAFCYRGAEVVISPLPNASAPAAQRNSQGAHLRQGVGGSDPTSCGVASG 179
Db 146 GGALLGVLYRGAELHQLPGLGGTNSA--GGPGAHLRRK---SPASGGPNCN--- 196
QY 180 WNPAILRALDPPKPRAGGSRSSRRSRRAKRFVSPRYVETLWVADESVMKFGADLE 239
Db 197 -----KAPLGSPSPRPR--RAKRFASLSRFVETLWVADEKMAAFHAGL 239
QY 240 HYLLTLTAARLYRHPSLINPINTVVKVLLLRDRDSGPKVTGNAALTIRNFCWAKKL 299
Db 240 RYLLTVMMAAKAFKHPISIRNPVSLVTVRLVILGSGEGPQVGPSPAAQTLRSCAWORGL 299
QY 300 NKVSKHPEYWDATLFTRLQDLCGATTCDTLGMADVGTWCPKRCSCSVIEDDGLPSAFTT 359

Db 300 NTPESDPDHFDTALFTRLQDLCGVSTCDTLMADVGTWCPKRCSCSVIEDDGLQSAFTA 359
QY 360 AHELGHVFNPHNDNPKYCEVFGKLI-RANHMSPTLIQIDRANPWSACSAIITDFLDSG 418
Db 360 AHELGHVFNPHNDNPKYCEVFGKLI-RANHMSPTLIQIDRANPWSACSAIITDFLDSG 419
QY 419 HGCDLLDQPSKPSLPLDPLGASVYLSQOCELAFGVGSKPCPYM-QYCTKLWCTGKAKQ 477
Db 420 YGCHLLDRPEALHLPVTFPGKYDADQCOLTFGDSRHCPLPPPCALWCSGHLNGH 479
QY 478 MYCQTRHFPWADGTSCGEGKCLKAGACVERHNLNKHRVD--GSWAKMDPYGCPSCRTCGG 535
Db 480 AMCQTKHSPWADGTSCGEGKCLKAGACVERHNLNKHRVD--GSWAKMDPYGCPSCRTCGG 539
QY 536 VOLARROCTNPTPANGKCYCEGVYKYRSCNLEPCPSSASGSKSPREOCQAFNGYNHSTN 595
Db 540 VOFSSRDCRTPRPNNGKCYCEGVYKYRSCNLEPCPSSASGSKSPREOCQAFNGYNHSTN 595
QY 596 RLTL---AVAWPKYSGVSPDRCKLICRANGTYFYVLPAPKVVDTLCSPDSTSCVQ 652
Db 596 LFKSPGMDWPRYTGVAPOQCKLTQARALGYVYVLEPRVVDGTPCSPDSSSVQV 655
QY 653 KCIKAGCDNLGSKRFDKCGVCGDGNKSKKVTG 687
Db 656 RCIHAGCDRIIGSKKFKDKCMVCGDGGSGCKSKQSG 690

RESULT 3

US-09-369-364A-9

; Sequence 9, Application US/09369364A

; Patent No. 6391610

; GENERAL INFORMATION:

; APPLICANT: Apte, Suneel

; APPLICANT: Hurskainen, Tiina L.

; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases

; FILE REFERENCE: 26473/4007/10-30-00

; CURRENT APPLICATION NUMBER: US/09/369.364A

; CURRENT FILING DATE: 1999-08-06

; NUMBER OF SEQ ID NOS: 31

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 9

; LENGTH: 905

; TYPE: PRT

; ORGANISM: Mus musculus ADAMTS-8

US-09-369-364A-9

Query Match 44.8%; Score 1679; DB 4; Length 905;
Best Local Similarity 46.7%; Pred. No. 3.4e-132;
Matches 344; Conservative 101; Mismatches 202; Indels 90; Gaps 18;

QY 2 LLLGLTLT-----APAGRTAGGFEPEREVVPRLDPDINGRRYWRGPDSDGGL 53
Db 13 LLLLLLQPPPLVCGAPAGPGTGA--QASELVVPTRL-----PGSASE--L 55
QY 54 IFQITAFQEDFYVLPDQAFAPAFSTHGLVPLQGLTGGSSDLRRRCFYSGDVNAEPDS 113
Db 56 AFHLSAFQGGFVRLAPDASFLAPFEKIERLG-GSSAAGGEPGLRCGFFSGTVNGERES 114
QY 114 FAVSLCGLRCGAFGYRGAEVVISPLPNASAPAAQRNSQGA-----HLLQRGVPGGP 166
Db 115 LAAMSCVAGWSGSLFAGEEFTIQP-----QGAGDSLDPHRLQRMG-PGOR 160
QY 167 SGPDSRSCGVASGWNPAIIRALDPPKPRAGGSRSS-----RRSG----- 208
Db 161 REDP---GLAAAEVFPPLQGLEWEVENGGQGGERSDNEEDKKQDEGLLKETEUSKV 216
QY 209 -----RAKRFVSPRYVETLWVADESVMKFGADLEHYLLTLTAARLYRHPISILN 260
Db 217 PPPFGSKTRSKRFVSEARFVETLWVADEMAAFYGTDLQNHILTVMSMAARIYKHPISIRN 276
QY 261 PINIVVVKVLLLRDRDSGPKVTGNAALTIRNFCWAKKLNVSKDKHPEYWDATLFTROD 320

```

Db 277 SVNLVVKVLIYKERWGPVSDNGGLTLRNFCSKQREFNKPSPRHEHYDTAILFTTRQN 336
QY 321 LCG-ATTCTDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEE 379
Db 337 FCGKGEQCDTLGMADVTGICDPDKSCSVIKDEGLQAAAYTLAHELGHVLSMPHDDSKPCVR 396
QY 380 VFGKLRANHMPTLIQIDRRANPWSACSAAIITDLSGHDGDCLLDOPSPISLPEDLPG 439
Db 397 LFGPMKYHMAFFIHNKTLNPSPCSAVYLTLLDDGHDGDCLLDAPTSLVLPDGLPG 456
QY 440 AS--YTLSSQCELAFGVSGKPCP---YMOYCTKLWCTGKAKQMWVCOTRH--FPWADGTS 492
Db 457 HSTLYELDQOQKIFGPDPRHCNPNTSVEDICVOLCARHSDSEPTICHTKNGSLLIWADGTP 516
QY 493 CGEGKLCAGACVERHNLN--KHRVDGSKWAKWDVPGPCSRCTCGGVOLARROCTNPTPAN 550
Db 517 CGPGHLCLOGSCVLKEDVENKAVVDGWPWRPWGQCSRTCGGIGTFSNRECDNMPQN 576
QY 551 GKGYCEGVRYKRYSCNLEPCPSSASGSKSFREEQCEAFNGYNHSTNRLTLAVAWVPKYSV 610
Db 577 GGRFCLGERVKYQSCNTEBEP--PNGKSFREQCEKYNAYNH--TDLDGNFLQWVPKYSV 633
QY 611 SPDRCKLICRANGTYFVLPAPKVVDGTLCPDSTSVCOGKCIKAGCDNLSKKRED 670
Db 634 SPDRCKLCFRARGSEFKVPEAKVIDGTLCPDSTSVCOGKCIKAGCDNLSKKRED 693
QY 671 KCGVCGGDNKSKCKVTG 687
Db 694 KCGVCGGKTACKRISG 710

```

RESULT 4
US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:

```

; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

```

```

Query Match 43.0%; Score 1611; DB 4; Length 930;
Best Local Similarity 47.3%; Pred. No. 1,7e-126;
Matches 305; Conservative 82; Mismatches 182; Indels 76; Gaps 7;

QY 90 GLFGSGSLLRRFCYSGDVNAEPDSPAVALSGLRGAFYGAEXVISPLNASAPAAQR 149
Db 121 GLSASSGHRGHCFYRGTVGDSRSLAVFDLCGLDGFFAVKHARYTLKPLRGSWAERY 180
QY 150 ----NSQAHLLQRRG-----VPGPSPGDPSTSRGVASGWNPAIRALDPY 191
Db 181 IYDGSRLIYVNRGFSFALPPRASCETPASFSG----- 217
QY 192 KPRAGFGESRRRRSG-----RAKRFVSIPRYVETLVVAD 227
Db 218 -POESPVSRRSRRRSALAPQLLDHSAFSPSGNAGPQTWRRRRRSISRQVLELLVAD 276
QY 228 ESMVKHGADLEHYLTATARLYRHSILNPNIVVVKVLLLRDSDGPKYTGNAAL 287
Db 277 SSMRWYGLQHYLLTMASIANRULYSHASIEHRLAVKVVVLTDKDTSLEVSNAAT 336
QY 288 TLRFCAWOKKLNKVDKHPYWDTAILFTRODLGATTCDTPLGMADVGTMCDPKRSCSV 347

```

```

Db 337 TLKNECKHQHQNOLGDDHEEHYDAAILFTREDLCGHHSCTDLGMADVGTICSPERSCAV 396
QY 348 IEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMPTLIQIDRRANPWSACS 407
Db 397 IEDDGLHAFTVAHEIGHLLGLSHDDSKFCBENFGTTEDKRLMSSILTSIDASKPWSKCT 456
QY 408 AAIITDFDLSHDGDCLLDQPSKPISLPEDLPASVYTLSSQCELAFGVSGKPCPYMOTYCTK 467
Db 457 SATITEFDLDDHGNCLLDLPKQILGPPELPGQYDATQCCLNLTGFGPEYSVCPGMDVCAR 516
QY 468 LMCTKAKQMWVCQRRHPFADGTSCGEGKLCAGACVER-----HNLNKHVRVDGSKAKW 522
Db 517 LMCAYVRGQMVCLTKLPAVEGTGCGKRYCLOGKCDKTKKKYKYSTSSH--GNGMSW 573
QY 523 DPGYPCSRCTCGGVOLARROCTNPTPANGKYCEGVRYKRYSCNLEPCPSSASGSKSPREE 582
Db 574 GPWGQCSRSGCGGVQFAYHCHNPNAPRNSGRYCTGKRAIYRSCSVTPCP--PNGKSFRE 631
QY 583 OCEAFNGYNHSTNRLTLAVAWVPKYSVSPDRCKLICRANGTYFVYVLPAPKVVDGTLCS 642
Db 632 QCEAKNGYQSDAKGVKTFEVWPKYAGVLPADVCKLTCAKGTGYVYVFSKPYTDGTECR 691
QY 643 PDSTSVCOGKCIKAGCDNLSKKREDKCGVCGGDNKSKCKVTG 687
Db 692 PYSNSVYVNRGRCVTRGCDIGSKLOYDKGCVCGGDNSSCTKIIG 736

```

RESULT 5
US-09-122-126B-15
; Sequence 15, Application US/09122126B
; Patent No. 6451575
; GENERAL INFORMATION:

```

; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGRECAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DMG909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-15

```

```

Query Match 42.4%; Score 1587.5; DB 4; Length 930;
Best Local Similarity 47.2%; Pred. No. 1.6e-124;
Matches 322; Conservative 82; Mismatches 209; Indels 69; Gaps 15;

QY 48 SGDOGLIFQITAFQEDFYHLHTPDQAFLAPAFSTEHLGVPLQGLTGSSDLR---RCFYS 104
Db 82 SGGGKGVLYVAGRRFLDLDERDGSVGTAGF-----VPAGG--GTSAPWRHSHCFYR 133
QY 105 GDVNAEPDSPAVALSGLRGAFYGAEXVISPL---PNASAPAAQRNSOGA---HLL 157
Db 134 GTVDASPRSLAVFDLCGLDGFFAVKHARYTLKPLRGPWAEKGRVYGGSGARILHY 193
QY 158 QRRG-----VPGGPS-----GDPTRSGVASGWNPAIR--RALDPYKP 193
Db 194 TREGFSFALPPRASCETPASTPEAHEHAPAHNSNPGRALAS---QLDQSALSP--- 246
QY 194 RRAFGES---RSRRSRAKRFVSIPIRVVETLVVADESVMKFFHCADLEHYLLTLLATAA 250
Db 247 -AGSGSPQTWRRRRRS-----ISRAQVELLLVADASMARLYGRGLQHYLLTLLASTAN 299
QY 251 RLYRHSILNPNIVVVKVLLLRDSDGPKYTGNAALTLRNFCAMQKKLNKVDKHPYEW 310
Db 300 RLYSHASIEHRLAVKVVVLGDKDSLEVSNAATTLKNFCQWQHQNOLGDDHEEHY 359
QY 311 DTAILFTRODLGATTCDTPLGMADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMP 370
Db 360 DAAILFTREDLCGHHSCTDLGMADVGTICSPERSCAVIEDDGLHAFTVAHEIGHLLGLS 419

```



```
Db 435 DTGCGCLLNEPSPYPLPVLPGILYNNKQXELIFGPGSQVCPYMMQCRLWCNNVN 494
QY 472 GKAKQWMCOTRHFHWADTSCGEGKCLKLGACVERHNLNKRVDGSGWAKWDPYPCSR 531
Db 495 GVHKG---CRTQHTWADTECEPGKHCKXGCVPK-EMDVPTVDSGWSWSPFTCSRT 550
QY 532 CGGQVLARROCTNTPANGKYGCGVRVYKRSCLNLEPCPSSASKSPREOCEAFNGYN 591
Db 551 CGGKIKTAIRECNRPEKNGGKYGVRMRKFKSCNTEPCLKQK---RDFRDEQCAHFDGKH 608
QY 592 HSTNRLTLAVAWPKYSGVSPDRCKLICKRANGTGYEVVLAPKVVDTGLCSPDSTSVCVQ 651
Db 609 FNINGLLPNVRWPKYSGILMKDRCKLCFRVAGNTAYTQLDRVLDGTPCGODTNDICVQ 668
QY 652 GKICRAGDGNLGSKKRDKCGVCGGDNKSKKVTG 687
Db 669 GLCROAGCDHVLNSKARRDKCGVCGGDNSSCKTVAG 704

RESULT 8
US-09-369-364A-15
; Sequence 15, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 874
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-9
US-09-369-364A-15

Query Match 37.2%; Score 1394.5; DB 4; Length 874;
Best Local Similarity 44.6%; Pred. No. 2.2e-108;
Matches 271; Conservative 86; Mismatches 202; Indels 49; Gaps 12;

QY 115 AAVSLCGGLGAFAGYGEYVISLPLNASAPAAQNSQGAHLRRGVPGSPGDPSTSRC 174
Db 3 AVISLCSGMMGTFRSHDGDYFIEPLQSVDEQEEQNKPHIYRHSTPQREPSTGKHAC 62
QY 175 GVASGNP-----AILRA-----LDPKPRAGGESRRR 205
Db 63 ATSELKNSHSDKIRMKRRKRNSLADVALLKSLGATKVLGYSNQT---NNTDRW 119
QY 206 RSGRAKRVSTPYVETLVVADESVMKFGADLEHLLTLATAARLYRHPISILNPNIV 265
Db 120 NKRKTRFLSYPREFVNMVADHRLVYHGANLQHYILTMSIVASYIKDSSIGNLNIV 179
QY 266 VVKVLLLRDSDGPKVTGNAALTIRNCAMQKLNKYSVD-KHPYWDTAILFTQDLCGA 324
Db 180 IYLVVHNEQEGPYINFAQTTLNFCOMQHSKNYLGGIQH----DTAVLVTRDICRA 235
QY 325 -TTCDTGLMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEYFGK 383
Db 236 QDKCDTLGLAELGICDPYRSCSISESGLSAFTIAHELGHVFNMPHDNSNCKEE-GV 294
QY 394 LRANHMSPTLIQIDRANPWSACSAIITDPLDSGHGDCILDDP-SKPISLPDLPGASY 442
Db 295 KSPQHVMAPTLNFYTNPMWMSKSRKVIETFDLTGYGECLLNEPASRTYPLPSQLPGLLY 354
QY 443 TLSQOCELAFGVGSKPCPYMYCTKLWCT---GKAKQWMCOTRHFHWADTSCGEGKLC 499
Db 355 NVNKOCELIFGPGSQVCPYMMQCRLWCNNVVDGAHKG---CKTQHTPWADGTCEPGKHC 411
QY 500 LKGAACVERHNLNKRVDGSGWAKWDPYPCSRCTCGGQVLARRCTNTPANGKYGCEVR 559
```

```
Db 412 KFGFCVPK-EMEGPAIDGSMGWSHFGTCSRTCGGKIKTAIRECNRPEKNGKYGVR 470
QY 560 VYBSCLNLEPCPSSASKSPREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPDRCKLI 619
Db 471 MKFKSCNTEPCKQK---RDFREOCEAHFDGKHFNINGLLPSPVRFWKYSGILMKDRCKLF 528
QY 620 CRANGTGYFYVLAPKVVDTGLCSPDSTSVCQVQKCIKAGCDGNLGSKKRDKCGVCGGDN 679
Db 529 CRVAGNTAYQLDRVLDGTPCGODTNDICVQGLCROAGCDHILNSKVRKDKCIGCGDN 588
QY 680 KSCKKVTG 687
Db 589 SSCKTVAG 596

RESULT 9
US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7

Query Match 28.8%; Score 1077; DB 4; Length 997;
Best Local Similarity 38.4%; Pred. No. 1.1e-81;
Matches 284; Conservative 82; Mismatches 266; Indels 108; Gaps 32;

QY 1 MLLGLILTLAFAGRTAGGPEPER---EVVVPRLDP-----DINGRR- 39
Db 18 LLLLCALAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 77
QY 40 ---YY---WRGPRDSGDQGLIFQITAFQEDFVYHLTPDAQFLAPAF--STEHLGVPLQGL 91
Db 78 APAFYELQYRGRE-----LRFNLITANQ-----HL-----LAPGFVSTRRRG---GL 116
QY 92 TGSSDLR---RCFYSGDVNAEPD---SFAAVSLCGGLGAFAGYGEYVISLPLNASE 144
Db 117 --GRAHIRAHTPACHLLGEVQ-DPELEGGGLAAISACDGLKGVFQLSNEDYFIEPLD--SA 171
QY 145 PAAQRNSQGAHLQRRGVP-----GGPSGDPSTRCGVASGWNPAILRALDPPKPRAGRG 199
Db 172 PAPGHAQ-PhVYKROAPERLAQRGDSAP-STCGV-----QYVPELESRRRW 219
QY 200 ESRSRRSRAKRF-----VSIPRYVETLVVADESVMKFGH-ADLEHYLLTLATAARLYR 254
Db 220 EQQQQRRRLRLRLHRSYSKEKWCETLVVADAKWVEYHGQPOVESYVLTIMWVAGLPH 279
QY 255 HPSILNPNIVVVKVLLLRDSDGPKVTGNAALTIRNFCAMQKLNKVKSDKHPEYWDTAI 314
Db 280 DPSIGNPIHITIVRLVLEDEEDLKITHADNTLKSFCCKWQKSNMKMDAHLPHHDTAI 339
QY 315 LFTQDLCGATTT---CDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHD 372
Db 340 LTRKDLCEAMNRPCTGLSLSHVAGMCQPHRSCNEDTGLPLAETVAHELGHSGFIQHD 399
QY 373 -NVKVEEYFGKLRANHMSPTLIQIDRANPWSACSAIITDPLDSGHGDCILDDPQSKP- 430
Db 400 GSGNDCFV-GK--RFFIMSPQLLDYAAFLTWRSRCSQYITFLDRGMLGLDDPPAKDI 456
QY 431 ISLPDLPGASYTLSQOCELAFGVGSKPCPYM-QYCTKLWCTGKAKQWMCOTRHFHWAD 489
```

```
Db 457 IDPSPVPGVLYDVSHQCRQLQYGAYSAFCEDMDNVCHTLCVSGT-----TCHSKLDAVD 512
QY 490 GTSCEGKGLCKGACVERHNLKRVDSWAKWDPYGPCSRTCGGVQLARRQCTNPTPA 549
Db 513 GTRCGENKWSLGECPV-VGFRPBAVDGSGWNSAWISCRSGMGVQSAERQCTOPTPK 571
QY 550 NGKPYCEGVRYKRSCLNLEPCSPSSASGKSFRFEOCEAFNGYNHSTNRLTLAVAWVPKYS 609
Db 572 YKGRYCVGERKFRCLNLOACP--AGRPSFRHVQCSHFDAWLYKGQHLT-----WVPVND 625
QY 610 VSPRDKKLCIRANGTYFYVVLAPKVDGTLCSL--DSTSVCVQOGKICAKGCDGNLGSKK 667
Db 626 VNP--CELHCRPANEXYFAKRLDACVDGTPCYQVRASRLCINGICKNVGCDFEIDSGA 682
QY 668 RFDKGVCGVGNKSKCKVTG 687
Db 683 MEDRCVCHGNGSTCHIVSG 702

RESULT 10
US-09-130-491-16
; Sequence 16, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodheart, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-130-491-16

Query Match 28.6%; Score 1072; DB 4; Length 551;
Best Local Similarity 58.7%; Pred. No. 1.2e-81;
Matches 185; Conservative 45; Mismatches 79; Indels 6; Gaps 4;

QY 377 CREVEGKLRANHMSPTLIQIDRANPWSACSAAIITDLSGHDCLLDQSPKPSLSPED 436
Db 1 CASLNGVSGDHLNLSLSDHSPWSPCSAYMVTSLDNGHGECLMDRKPQNPILPSPD 60

QY 437 LPGASYTLSSQCELAFLGKVGSKPCP-YMQYCTKLWCTKAKGQWVQTRFHPWADGTSCEG 495
Db 61 LPTLYDANRQCQFTFGESKHCDAASTCTTLWCTGTSGLLVCQTKHFPWADGTSCEG 120

QY 496 GKLCILKACVERHNLNKH---RVDGSAKWDPYGPCSRTCGGVQLARRQCTNPTPANG 552
Db 121 GKWCVSGKCVNKTDM-KHFAFPVHSGWPGWPGWDCSRTCGGVQVYTMRECDNPVPKNG 179

QY 553 KYCEGVRYKRSCLNLEPCSPSSASGKSFRFEOCEAFNGYNHSTNRLTLAVAWVPKYSVSP 612
Db 180 KYCEGVRYKRSCLNLEPCSP--NKTFRFEOCEAHNFESGAFNGEFTVWTPRYAGVSP 238

QY 613 RDKCKLCIRANGTYFYVVLAPKVDGTLCSL--DSTSVCVQOGKICAKGCDGNLGSKKRFDK 672
Db 239 KDRCKLTCEAKGIGYFVLQPKVVDGTPCSPDSTSVQVQCVKAGCDRIIDSKKRFDKC 298

QY 673 GYCGGDNKSKCKVTG 687
Db 299 GYCGGNGSTCKRMMSG 313

RESULT 11
```

```
US-09-369-364A-22
; Sequence 22, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-5
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (99)
; OTHER INFORMATION: Xaa = M
US-09-369-364A-22

Query Match 26.8%; Score 1004; DB 4; Length 518;
Best Local Similarity 52.9%; Pred. No. 5.7e-76;
Matches 174; Conservative 42; Mismatches 103; Indels 10; Gaps 3;

QY 364 GHVFNNPDNPKVCEEVFGKLRANHMSPTLIQIDRANPWSACSAAIITDLSGHDCL 423
Db 1 GHLLGLSHDDSKFCEETFGSTEDKRLMSSILTSIDASKPWSKCTSATITEFLDDGHGNC 60

QY 424 LDQSPKPSLSPEDLPGASYTLSSQCELAFLGKVGSKPCPYMQYCTKLWCTKAKGQWVQTR 483
Db 61 LDLPRKQILGPEELPGQTYDATQCNTLTFGPDYSVCPGXDVCAWLCAVVRQGMVCLTK 120

QY 484 HFPWADGTSCEGKLCILKACVER-----HNLNKHRYVDGSAKWDPYGPCSRTCGGVQL 538
Db 121 KLPVVEGTPCGKGRICLOGKCVDTKKKYSTSSH---GNWGSWGSWGQCSRSCTGGVQ 177

QY 539 ARQCTNPTPANGKYCEGVRYKRSCLNLEPCSPSSASGKSFRFEOCEAFNGYNHSTNRLT 598
Db 178 AYRHCNNPAPRNNRGYCTGKRAIYHSCSLMPCP--PNKSFREHCEQCEAKNGYOSDAKGV 235

QY 599 LAVAWVPKYSVSPRDKKLCIRANGTYFYVVLAPKVDGTLCSL--DSTSVCVQOGKICAK 658
Db 236 IFVWVVPYAGVLPADVCLTKRAKGTGYVYVFPSPKVTDTGTECRPYNSVCVRGKCVRTG 295

QY 659 CDGNLGSKKRFDKGVCGGDNKSKCKVTG 687
Db 296 CDGIIGSKLQYDKGVCGGDNSSCTKIYG 324

RESULT 12
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Collige, Alain
; APPLICANT: Lapiere, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/491,522
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886,333
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8389-0060-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-491-522-5

Query Match 26.1%; Score 976; DB 4; Length 1211;
Best Local Similarity 35.5%; Pred. No. 4e-73;
Matches 260; Conservative 93; Mismatches 282; Indels 98; Gaps 31;
QY 9 LAFAGTAG--GFPEPEVVVPIRLDPD-----INGRR-----YYW 42
Db LAAADPPGPIGHAERILAVPVRTDAQGRLVSHVSAATSRAGVARRAARVPTSEFP 98
QY 43 RGPEDSGDQGLIFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCF 102
Db GNEEEPESHLYFNVTVGRDLRLRNLARLVAPGATMEMQGE--KGTTRVEPLLGSL 156
QY 103 YSGDWN--AEPDSFAAVSLCGGLRGAFYGAEYVISPPLNASAPAAQNSQG-AHLIQR 159
Db YVGDVAGLAEASS-VALSNCGLAGLIRMEEEFFIEPLEKGL--AAQEAQGRVHVYR 213
QY 160 RGVPGSGDPTSRGCVASGNPAILRALDPYKPRAGFGESRRRRSGRAKRVSTIPRY 219
Db RPTSPPLGGPQA---LDTG---ASLSDLSL--RALGVLEEHAANSRRRRRAHAADDY 266
QY 220 VE-TLVVADESNVFKHGAD-LEHYLLTLTAAALYRHPSTILNPINIVVKKVLLLRDSD 277
Db NEIVLLGVDDSVVQPHGKEHVQKYLTLNINVEIYHDESAGAHNVVLYRIILLSTGKS 326
QY 278 GPKV-TGNAALTLRNFCAWQKLNKVS DKHPYWDTAILFTRODLGCATTCDTLGMADVG 336
Db MSLEIGNPSQSLENVCWAYLQOKPDTGHDYHDHAIFTLRQDF-GPSGMQ--GYAPVT 383
QY 337 TMCDDPKRSCVIEDDGLPSAFTTAHELGHVFNMPHDNV-KVCEVEFGKLRANHMMSPTLI 395
Db GMCHPVRCSTLNHDEGFSFAVFAHEGTGVLGMEHDGQGNRCGD---EVRLGSTMAPLVQ 440
QY 396 QIDRANPWSAGSAIITDFLDGSHGDCLLDQD---SKPISLPEDLPGASTYTLSSQCELA 452
Db AAFHFRHWSRCSQELSRYLHS--YDCLLDDPPFAHDWP-ALPQ-LPGLHYSMNEQCRFDF 496
QY 453 GVSGKPCPYMOY---CTKLWCTGAKGOMVQTRHFPWADGTSCEGKCLKLGACV---- 505
Db GILGYMMCTAFRTFDPCKQLWCS-HPDNPYFCKTKKGPPLDGTMCAPGKHCFKGHCILWTP 555
QY 506 ---EHNHLNKRHVGSWAKWDPYGPCSRTCGGVQLARRQCTNPTPANGKYCEGVRYKY 562
Db DILKRD-----DGSWGAWSPFSCSRTCTGTVKFTROCDNPHANGGRTCSGLAYDF 607
QY 563 RSCNLEPCPSSAGSKSFREOEAFNGY-----NHSTNRLTLAVAVVPKYSGVSPRDK 615
Db QLCRSQDCPDSLA--DFREEOCRWDLYFEHGDQHH-----WLP-HEHRDAKER 654

QY 616 CKLICRANGTGYFYVLAPKVVVDGTLCS-POSTSVCVGOGKCIKACDGNLGSKRFRKCGV 674
Db 655 CHLYCESRETGEVYVSMKRMVHDGTRCSYKDAFSLQVRGDCRKGVCDCGVIGSSQEDKCGV 714
QY 675 CGGDNKSKCKVKG 687
Db 715 CGGDNCHCKVKG 727
RESULT 13
US-09-930-872-4
Sequence 4, Application US/09930872
Patent No. 6448388
GENERAL INFORMATION:
APPLICANT: Friidie, Carl Johan
TITLE OF INVENTION: No. 6448388el Human Proteases and Polynucleotides Encoding the
FILE REFERENCE: LEX-0219-USA
CURRENT APPLICATION NUMBER: US/09/930,872
CURRENT FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: US 60/225,852
PRIOR FILING DATE: 2000-08-16
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1224
TYPE: PRT
ORGANISM: homo sapiens
US-09-930-872-4

Query Match 25.9%; Score 968.5; DB 4; Length 1224;
Best Local Similarity 35.1%; Pred. No. 1.7e-72;
Matches 243; Conservative 86; Mismatches 261; Indels 103; Gaps 24;
QY 51 QGLIFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSS-----DLRRCFY 103
Db ESLHLRLKGSRHDFHVDLRTSSSLVAPGFIVQTL-----GKTGTSVQTLPPEDF--CFY 147
QY 104 SGDVAEPDFAAVSLCGGLRGAFYGAEYVISPPLNASAPAAQNSQG---AHLQRR 160
Db 148 QGSLRSHRNSVALSTCOGLSGMIRTEADYFLRPLSHLSWKLGRAAQGSFSPSHLYKR 207
QY 161 G-----VPGSGDPTSR-----CGVASGNWPA-----ILRA 187
Db STEPHAPGASEVLVTSRTWELAHQPLHSSDLRLGLPKQHFGRKKYKMPQPPKEDLFI 267
QY 188 LDYKPRRAGFGESRRRRSGRAKRVSTIPRYVETLVVADESVMKPHG-ADLEHYLLTL 246
Db PDEYK---SCLRHKSLRSHRNEEL-----NVETLVVVDKKMQNHENITTYVLTIL 319
QY 247 ATAARLYRHPSTILNPINIVVKKVLLLRDRDSGPKVTGNALTLRNFCAWQKL-NKVSOK 305
Db 320 NMYSAFLKDGCTIGTNINIAIVGLILLEDEQPLVISHHADHTLSSEFCQWGLMGKDGR 379
QY 306 HPEYWDTAILFTRODLGC--ATTCDTLGMADVGTCMDPKRSCSVIEDDGLPSAFTTAHEL 363
Db 380 H----DHAILLTGDLICSMKNPECDTLGFAPISGMCSKYRSCTINEDTGLGLAFTAHES 435
QY 364 GHVFNMPHDNV-KVCEVEFGKLRANHMMSPTLIQIDRANPWSAGSAIITDFLDGSHGDC 422
Db 436 GHNFIMHIDGEGNCKKSEG-----NIMSPTLAGRNVFWSWSPCSRYQLKHLFLSTAQAIC 490
QY 423 LLDQPSKPI--SLPEDLPGASTYTLSSQCELAFGVSKPCPY---MOYCTKLWC--TGKA 474
Db 491 LADQP-KPVKEYKYPEKLPGLDYDANTQCKWQFGEKAKLCMLDFKDKIDCKALWCHRIGR 549
QY 475 KGOMVQTRHFPWADGTSCEGKCLKLGACVERHNLNKRHVGSWAKWDPYGPCSRTCGG 534
Db 550 -----CETKFMPAAGETICGHDWVGCGQCVKYGDGPKPTHGHWSDWSWSPCSRTCGG 604
QY 535 GVOLARRQCTNPTPANGKYCEGVRYKRSCLNEPCPSSAGSKSFREOC-----EAFNG 589
Db 605 GVSHRSLCTNPKPSHGKFCGEGSTRTTLKLCNSQKCPRDS--VDFRAAQAEHNSRRFRG 662

```
QY 590 YNHSTNRLTLAVAVPKYSGVSPROCKLILCRANGTGYFYVLAPKVVDTGLCSFDSVSVC 649
Db 663 RHYK-----WKP-YTQVEDDGLCKLYCIAEGDFDFSLNKNKVDGTPCSEDSRNVC 712
QY 650 VOGKIKAGCDNGLGSKRRPKDCVCGCGDNKSC 682
Db 713 IDGICERVGCNDVLSDAVEDVCVCGNGNSAC 745

RESULT 14
US-09-491-522-11
; Sequence 11, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Collige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEFAX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-11

Query Match 25.2%; Score 944.5; DB 4; Length 1205;
Best Local Similarity 33.5%; Pred. No. 1.7e-70;
Matches 256; Conservative 96; Mismatches 276; Indels 137; Gaps 32;

QY 1 MLLGILFLAFAGRTAG-----GFPEREVVVPIRL----- 32
Db 16 LLLLLLLPLADARLAAADDPGPGQGHGAERILAVPVRTDAQGLVSHVSAATAPAG 75
QY 33 -----PDINGRIRYWRGPDSDGDLFIQITAFQEDFVHLHTPDAQFLAPAFST 81
Db 76 VTRRAAPAQIPGLSG-----GSEDPGGRLFNYTVFGDRDLHLRFLNARLVAGATV 129
QY 82 EHLGVLQGLTGGSDLRRCFYSGDV-NAEPDSPAAYSLCGGLRGAFYRGAEYVISPPL 140
Db 130 EQWQE--SGATRVPEPLTCLYVGDVAGLAESSVALSNCDLAGLIRMBEEFFIEPLE 187
QY 141 NASAPAAQRNSQG-AHLLORRGVPGSGDPTSR-----CGVASGWNPAIRAL 188
```

```
Db 188 KGL--AAKEAQGRVHVYHRT-----TSRFPLGFGQALDTGISADSLDSLRAL 237
QY 189 DPYKRRAGFGESSRRSRGAKRFVPIRY-VETLVVADESVMVKFHGAD-LEHYLLTLL 246
Db 238 -----GVLEERVNSRRMRRAADDDNYIEVLGVDDSVVQFHGTEHVQKYLTLTM 289
QY 247 ATAARLYRHPHSILNPINIVVYVLLLRDROSGPKV-TGNAALTLRNFCAMOKKLNVSDK 305
Db 290 NIVNEIYHDESIGAHINVLRIILLYSGKSMLEIGNFQSLNENVCWAYLQKRPDTD 349
QY 306 HPEYNDTAILFTRODLGATCTDLGMADYGTMDPKRSCSVTDDGLPSAFTTAHELGH 365
Db 350 HDEYHDHAIFLTRQDF-GPSGMQ--GYAPVTGMCHPVRSCTLNHEDGFSFAFVAHETGH 406
QY 366 VENMPHDNV-KVCEEVFGKLRANHMSPTLIQIDRANPWSACSAAITDPLDSGHGDCLL 424
Db 407 VLGMHHDGQGNRCGD---EVLGSIAPLVQAAAFHRFHWRSRCSQOQLSRYLHS--YDCLR 461
QY 425 DQP---SKPISLPEDLPASVYLLSQOCELAFGVGSKPCPYMQY---CTKLWCTGKAKGOM 478
Db 462 DDPETHDWP-ALPQ-LPGLHYSMNEQCRDFDGLGYMCTAFRTFDPCKQLWCS-HPDNPY 518
QY 479 VCQTRH-FPWADGTSCGEGKCLKAGAC-----VERHNLNKHVRVDSHAKWDPYGPCSR 530
Db 519 FCKTKKGEP-LDGTMCAPGKHCFKGHCTWLTFTDLKR-----DGNMANSFPGCSR 569
QY 531 TCGGVOLARROCTNPTPANGGKYCEGVVRVYKSCNLEPCSPSSAGSKSPFEECEAFNGY 590
Db 570 TCGTGVKTRQCDNPHANGRTCSGLAYDFOLCNSQDCPDALA--DFREEQCRWDLY 627
QY 591 -----NHSTNRLTLAVAVPKYSGVSPDKCKLILCRANGTGYFYVLAPKVVDTGLCS- 642
Db 628 FEHGDAQRH-----WLP-HEHRDAKERCHLYCESKETGEVYMKRMVMDHGTSCSY 676
QY 643 PDSTSVCOVKICIRAGDGNLGSKKRFDKGVCGGDNKSKCKVTG 687
Db 677 KDAFSLVGRDCRVGCGDGVIGSSKQEDKGVCGGDNHCKVKVG 721

RESULT 15
US-09-369-364A-17
; Sequence 17, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Aptec, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-10
US-09-369-364A-17

Query Match 24.9%; Score 930.5; DB 4; Length 1081;
Best Local Similarity 35.3%; Pred. No. 2.2e-69;
Matches 232; Conservative 82; Mismatches 276; Indels 67; Gaps 22;

QY 64 FYLHLPDAQFLAPAFSTHGLGVPLQGLTGGSDLRRCFYSGDVNAE-PDSFAAVSLCGG 122
Db 65 FLLNTRSSRLLAGRVSEYV--TREGIAWRARPHCLYAGHLQQAQSSHVAITCGG 122
QY 123 LRGAFGYRGAEYVISPPLNAPASAPAAQRNSQGAHLQRRGVPGSGDPTSRCCV--ASGW 180
Db 123 LHGLIVADEEYLIETPLHGGPKGSRSPESGPHCVYKRSLRHPHLD--TACGVREKWP 180
QY 181 --NPAIRALDPYKRRAGFGESSRRSRGAKRFVSIPIRYVETLVVADESVMVKFHG-AD 237
```

```

Db 181 KGRPWLLTKP--PPARPLGNETERGPG-LKRSYSRERYVETMDVADKMMVAYHGRD 237
QY 238 LEHYLLTLLATAARLYRHPSTILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWOK 297
Db 238 VEQYVLAIMNTIVAKLFQDSSLGSTVNILVTRILLTEDOPTLEITHAGKSLDSECKWOK 297
QY 298 KL-----NKVSKHPEYWDTALLFTRODL--GATTCDTLGMADVGTMCDPKRSCSVI 348
Db 298 SIVNHSHGHNAPENGVAHDATVLTITRYDICIYKNKPCGTGLAR-WAECYSAREAAAS 356
QY 349 EDDGLPSAFTTAHELGHVFNPHDNVKEVEFG-----KLRAHHMSPTLIQIDRANP-- 402
Db 357 MRTLAAATSVHHCHEIGHTFGMNHGVDGNSCGARGQDPAKLMAAHITMKT-----NPFV 409
QY 403 WSACSAALITDFLDSHGDCILLDOPSK-PISLPEDLPGASYTILSQOCELAFAFGVSKPCPY 461
Db 410 WSSCNRDYITSLDGLGLCLNRPQDFVYPTVAPGQAYDADEQCRFQHGKVSQCKY 469
QY 462 MOYCTKLWCTGKAGQWVCQTRHPHPWADGTSCEGKL-----CLKGACVERHNLNKHVRDG 517
Db 470 GEVCSLWCLSKSNR---CITNSIPAAEGTICQHTTIDKGWCYKRVCP--FGSRPEGVDG 525
QY 518 SWAKWDYPGCSRTCGGGOVLARQCTNPTPANGGKYCEGVRYKYRSCNLEPCPSSASGK 577
Db 526 ANGPTWPGDCSRTCGGVSRRHCDSPRTTGGKYCLGERRRHRSCTNTDDCPPGS--Q 583
QY 578 SFREQCEAFNG-----YNHSTNRLTLAVAWVPKYSGVSPRDKCKLICRANGTGYFYV 630
Db 584 DPREVQCAEFDSIIPRGKFKWKTYR-----GGGVK---ACSLTSLAEGFNFTYE 630
QY 631 LAPKVVDTLCSPDSTSVQVQCKIKAGCDGNLGSKKRFDKCGVCGDKNKCKKVTG 687
Db 631 RAAAVVDGTPCRPDTVICVSGECKHVGCORVLGSLREDKCRVCGDGSACETIEG 687

```

Search completed: April 29, 2003, 17:11:55
Job time : 15.1439 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:12:21 ; Search time 15.2713 seconds
(without alignments)
3604.758 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_687

Perfect score: 3744

Sequence: 1 MLLGLTLTAFAGRTAGGFE.....RFDKCGVCGDNKCKKVTG 687

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB_Pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB_Pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB_Pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB_Pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB_Pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB_Pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/FCIUS_PUBCOMB_Pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB_Pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB_Pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB_Pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB_Pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB_Pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB_Pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB_Pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3736	99.8	950	10	US-09-965-631-4
2	3707.5	99.0	823	9	US-10-163-316-2
3	1960.5	52.4	967	12	US-10-105-929-2
4	1947.5	52.0	950	10	US-09-321-987B-4
5	1947.5	52.0	968	9	US-10-163-316-7
6	1913	51.1	367	10	US-09-965-631-6
7	1751	46.8	727	9	US-10-097-597-1
8	1751	46.8	727	9	US-10-097-597-1
9	1738	46.4	727	10	US-09-445-023A-1
10	1738	46.4	727	9	US-10-097-597-12
11	1738	46.4	727	9	US-10-097-580-12
12	1738	46.4	727	10	US-09-445-023A-12
13	1688.5	45.1	837	9	US-10-174-590-352
14	1688.5	45.1	837	9	US-10-176-758-352
15	1688.5	45.1	837	9	US-10-175-737-352
16	1688.5	45.1	837	9	US-10-173-706-352
17	1688.5	45.1	837	9	US-10-173-706-352
18	1688.5	45.1	837	9	US-10-175-738-352
19	1688.5	45.1	837	9	US-10-175-752-352
20	1688.5	45.1	837	9	US-10-176-482-352

20	1688.5	45.1	837	9	US-10-176-757-352	Sequence 352, App
21	1688.5	45.1	837	9	US-10-176-913-352	Sequence 352, App
22	1688.5	45.1	837	9	US-10-180-552-352	Sequence 352, App
23	1688.5	45.1	837	9	US-10-180-557-352	Sequence 352, App
24	1688.5	45.1	837	9	US-10-173-700-352	Sequence 352, App
25	1688.5	45.1	837	9	US-10-174-572-352	Sequence 352, App
26	1688.5	45.1	837	9	US-10-174-579-352	Sequence 352, App
27	1688.5	45.1	837	9	US-10-174-582-352	Sequence 352, App
28	1688.5	45.1	837	9	US-10-174-588-352	Sequence 352, App
29	1688.5	45.1	837	9	US-10-175-739-352	Sequence 352, App
30	1688.5	45.1	837	9	US-10-175-740-352	Sequence 352, App
31	1688.5	45.1	837	9	US-10-175-743-352	Sequence 352, App
32	1688.5	45.1	837	9	US-10-176-488-352	Sequence 352, App
33	1688.5	45.1	837	9	US-10-176-492-352	Sequence 352, App
34	1688.5	45.1	837	9	US-10-176-747-352	Sequence 352, App
35	1688.5	45.1	837	9	US-10-176-750-352	Sequence 352, App
36	1688.5	45.1	837	9	US-10-176-985-352	Sequence 352, App
37	1688.5	45.1	837	9	US-10-176-987-352	Sequence 352, App
38	1688.5	45.1	837	9	US-10-176-991-352	Sequence 352, App
39	1688.5	45.1	837	9	US-10-176-993-352	Sequence 352, App
40	1688.5	45.1	837	9	US-10-176-993-352	Sequence 352, App
41	1688.5	45.1	837	9	US-10-184-658-352	Sequence 352, App
42	1688.5	45.1	837	9	US-10-173-695-352	Sequence 352, App
43	1688.5	45.1	837	9	US-10-173-697-352	Sequence 352, App
44	1688.5	45.1	837	9	US-10-173-705-352	Sequence 352, App
45	1688.5	45.1	837	9	US-10-174-576-352	Sequence 352, App

ALIGNMENTS

RESULT 1

US-09-965-631-4

; Sequence 4, Application US/09965631

; Patent No. US20020115842A1

; GENERAL INFORMATION:

; APPLICANT: Friddle, Carl Johan

; APPLICANT: Hilbun, Erin

; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco

; FILE REFERENCE: LEX-0241-USA

; CURRENT APPLICATION NUMBER: US/09/965.631

; PRIOR FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: US 60/236,689

; PRIOR FILING DATE: 2000-09-29

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: fastseq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 950

; TYPE: PRT

; ORGANISM: homo sapiens

US-09-965-631-4

Query Match	99.8%	Score	3736;	DB	10;	Length	950;
Best Local Similarity	99.9%	Pred. No.	1.8e-276;				
Matches	686;	Conservative	0;	Mismatches	1;	Indels	0;
Gaps	0;						

Qy	1	MLLGLTLTAFAGRTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDGQGLIFQITAF	60
Db	1	MLLGLTLTAFAGRTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDGQGLIFQITAF	60
Qy	61	QEDFYHLTPDAQFLAFSTFHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDFAAAYSIC	120
Db	61	QEDFYHLTPDAQFLAFSTFHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDFAAAYSIC	120
Qy	121	GGLGAFGYRGAEVVISPLNAPAAQRNSQGAHLORGVPGSPGSDPTSRGCVASGW	180
Db	121	GGLGAFGYRGAEVVISPLNAPAAQRNSQGAHLORGVPGSPGSDPTSRGCVASGW	180
Qy	181	NPAILRALDPYKPRRAGFGESRRSRRAKRFVSIPIRYVETLVVADESVMKFGADLEH	240
Db	181	NPAILRALDPYKPRRAGFGESRRSRRAKRFVSIPIRYVETLVVADESVMKFGADLEH	240
Qy	241	YLLTLLATAARLYRHPISILLPINIVVVKVLLLRDRDGGPKVTGNAAALTLRNCAWOKKLN	300

Db 241 YLLTLLATAARLYRHPHSILNPNINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWOKKLN 300
Y 301 KVSOKHPEYWDTAILFTRODLGATTCDTLGMADVGTMCDPKRSCSVIEDDGLPSAFTTA 360
Db 301 KVSOKHPEYWDTAILFTRODLGATTCDTLGMADVGTMCDPKRSCSVIEDDGLPSAFTTA 360
Y 361 HELGHVFNPHDNDVNVKCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSHG 420
Db 361 HELGHVFNPHDNDVNVKCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSHG 420
Y 421 DCLLDQPSKPISLPEDLPASAYTLSQOCELAFGVSKPCPYMOYCTKLWCTGKAKQWVC 480
Db 421 DCLLDQPSKPISLPEDLPASAYTLSQOCELAFGVSKPCPYMOYCTKLWCTGKAKQWVC 480
Y 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKRH 540
Db 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKRH 540
Y 541 RQCTNPTPANGKYCEGVRVYKRSCLNLEPCPSSASGSKFRECEAFNGYNHSTNRLTLA 600
Db 541 RQCTNPTPANGKYCEGVRVYKRSCLNLEPCPSSASGSKFRECEAFNGYNHSTNRLTLA 600
Y 601 VAWPKYSGVSPRDKCKLICRANGTYFYVYIAPKVVDGTLCSPDSTSVCVQKCIKAGCD 660
Db 601 VAWPKYSGVSPRDKCKLICRANGTYFYVYIAPKVVDGTLCSPDSTSVCVQKCIKAGCD 660
Y 661 GNLGSKRRFDKCGVGGDNKSKKVTG 687
Db 661 GNLGSKRRFDKCGVGGDNKSKKVTG 687

RESULT 2
US-10-163-316-2
; Sequence 2, Application US/10163316
; Patent No. US2002019703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MP101-025PIRNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; PRIOR FILING DATE: 2002-06-05
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2

Query Match 99.08; Score 3707.5; DB 9; Length 823;
Best Local Similarity 96.58; Pred. No. 2.3e-274;
Matches 685; Conservative 0; Mismatches 27; Indels 23; Gaps 1;

Y 1 MLLGILTLAFAGTAGGPEPEVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
Db 1 MLLGILTLAFAGTAGGPEPEVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
Y 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDFAAAYSLC 120
Db 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDFAAAYSLC 120
Y 121 GGLRGAFAFYRGAEEYISPLPNASAPAAQNSQGAHLQRRGVPGSGDPTSRGCVASGW 180
Db 121 GGLRGAFAFYRGAEEYISPLPNASAPAAQNSQGAHLQRRGVPGSGDPTSRGCVASGW 180
Y 181 NPAILRALDPYKPRRAGGEGSRRRSRRAKRFVSIPIRYVETLVVADESVMYKFGADLEH 240
Db 181 NPAILRALDPYKPRRAGGEGSRRRSRRAKRFVSIPIRYVETLVVADESVMYKFGADLEH 240

Y 241 YLLTLLATAARLYRHPHSILNPNINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWOKKLN 300
Db 241 YLLTLLATAARLYRHPHSILNPNINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWOKKLN 300
Y 301 KVSOKHPEYWDTAILFTRODLGATTCDTLGMADVGTMCDPKRSCSVIEDDGLPSAFTTA 360
Db 301 KVSOKHPEYWDTAILFTRODLGATTCDTLGMADVGTMCDPKRSCSVIEDDGLPSAFTTA 360
Y 361 HELGHVFNPHDNDVNVKCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSHG 420
Db 361 HELGHVFNPHDNDVNVKCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSHG 420
Y 421 DCLLDQPSKPISLPEDLPASAYTLSQOCELAFGVSKPCPYMOYCTKLWCTGKAKQWVC 480
Db 421 DCLLDQPSKPISLPEDLPASAYTLSQOCELAFGVSKPCPYMOYCTKLWCTGKAKQWVC 480
Y 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKRH 540
Db 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKRH 540
Y 541 SWAKWDPYGPCSRTCGGGVQLARRQCTNPTPANGKYCEGVRVYKRSCLNLEPCPSSASGK 577
Db 541 SWAKWDPYGPCSRTCGGGVQLARRQCTNPTPANGKYCEGVRVYKRSCLNLEPCPSSASGK 577
Y 578 SPRECEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYFYVYIAPKVVD 637
Db 601 SPRECEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYFYVYIAPKVVD 660
Y 638 GTLCSPDSTSVCVQKCIKAGCDGNLGSKRFRDKCGVGGDNKSKKVTG 687
Db 661 GTLCSPDSTSVCVQKCIKAGCDGNLGSKRFRDKCGVGGDNKSKKVTG 710

RESULT 3
US-10-105-929-2
; Sequence 2, Application US/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: Tango-71, Tango-73, Tango-74, Tango-76, and Tango-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-105-929-2

Query Match 52.4%; Score 1960.5; DB 12; Length 967;
Best Local Similarity 53.6%; Pred. No. 3.7e-141;
Matches 388; Conservative 91; Mismatches 176; Indels 67; Gaps 17;

Y 1 MLLGILTLAFAGTAGGPEPEVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
Db 36 LLLAALAVSALGRPSDEELVVP-ELE-----RAP---GHGTTLRLHAF 81
Y 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGSSDLRRCFCYSGDVNAEPDFA 114
Db 82 DQQLDLRLPDSFLAPAGFTLQNVGRKSGSETPLP-----ETDLAHCFTSGTVNGDPSSA 136
Y 115 AAVSLCGLRGAFYRGAEEYISPLPNAS---APAAQRNSQGA-----HLQ---RRGVPG 165
Db 137 AALSCEVRGAFYLLGEAYFIQPLPAASERLATAPEKPPAPLQFHLRLRRNRQDVG 196

```
QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPKPRRAGGESRRRS 207
Db 197 TCGVDDPRPTGKAETDEDEGEDEGPQWS-----PODPALQGVGP-TGTGS 247
QY 208 GRAKREVSIPRYVETLVVADESVMKFGADLEHYLLTLTATAARLYRHPISILNPVWV 267
Db 248 IKRKREVSIRYVETLVVADESVMKFGADLEHYLLTLTATAARLYRHPISILNPVWV 307
QY 268 KVLRLDRDSGPKVTGNAALTLRNFCAWKKLNKVSOKHPEYWDYTAIFTRQDLCGATTC 327
Db 308 KILVTHDSQKGPVETVSNAAALTLRNFCAWKKLNKVSOKHPEYWDYTAIFTRQDLCGATTC 367
QY 328 DTGLGADYGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNHNDKVCVEEVEGKLRAN 387
Db 368 DTGLGADYGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNHNDKVCVEEVEGKLRAN 427
QY 388 HMASLTLQIDRANPWSACSAIITDFTDLSHGDCLLDOPSKPISLPELPGASYTLSQ 447
Db 428 HMASLTLQIDRANPWSACSAIITDFTDLSHGDCLLDOPSKPISLPELPGASYTLSQ 487
QY 448 CELAFGVGSKPCP--YMOYCTKLWCTGKAKGOMVCTRHFHWADGTSCEGKCLKAGACVE 506
Db 488 COFTFEGDSKPCPDAASTCTTLWCTGSGVLLVCOTKHFHWADGTSCEGKCLKAGACVE 547
QY 507 RHNLNKH---RVDGSWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKGYCEGVRYKR 563
Db 548 KTD-RKHFDTPFHSGWGMWGPWGDSCRTCGGGVQYTWRECDNPVYKNGKGYCEGVRYKR 606
QY 564 SCNLEPCPSSASGSKFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCLICRAN 623
Db 607 SCNLEPCPSSASGSKFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCLICRAN 665
QY 624 GTGYFYVLAPKVDGTLCSPDSTSVCVQGCICAKGDCNGLSKRFDKCGVCGDNKSK 683
Db 666 GIGYFYVLAPKVDGTLCSPDSTSVCVQGCICAKGDCNGLSKRFDKCGVCGDNKSK 725
QY 684 KVTG 687
Db 726 KISG 729

RESULT 4
US-09-321-987B-4
; Sequence 4, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Kimble, Judith E
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296.95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/129,023
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Murine
US-09-321-987B-4
Query Match 52.0%; Score 1947.5; DB 10; Length 950;
Best Local Similarity 52.3%; Pred. No. 3.5e-140;
Matches 381; Conservative 95; Mismatches 176; Indels 77; Gaps 16;
```

```
QY 1 MLLGILTLAPAGRTAGG--FEPEREVVPVIRLDPDINGRRYWRGP-EDSGDQGLIFOI 57
Db 20 LLLASITMLLCARGHGRTEDEELVLP-SLE-----RAPGHDSTTRL--RL 66
```

```
QY 58 TATQEDFYHLTPDAQFLAPAFSTEHGLVPLQGLTGS-----SDLRRCFYSGDVNA 109
Db 67 DAFGQQLHLKLQDPSGLFAPGFTQTV-----GRSPGSEAQHLDPDGLAHCFYSGTVNG 121
QY 110 EPDSFAVSLCGGLRGAFGYRGAEEYVLSPLPNAS---APAAQRNSOGA---HLQRRG 161
Db 122 DPGSAALSLCEGVRGAFYLOGEEFTIQAPGVATRELAPAVPEESSARPQFHILRRR- 180
QY 162 VPGGSGDPTSRGCVASGNPAILRALDPYKPRRAGFGESRRRR----- 206
Db 181 -----RRSGGAKCGVMD-----DETLPDTSRSPESQNTRNQWPYRDPDPODAGRP 226
QY 207 -----SGRAKREVSIPRYVETLVVADESVMKFGADLEHYLLTLTATAARLYRHPISILNP 262
Db 227 SGPGSIRKKRFVSSPRYVETLVVADESVMKFGADLEHYLLTLTATAARLYRHPISILNP 286
QY 263 NIYVVVLLRLDRDSGPKVTGNAALTLRNFCAWKKLNKVSOKHPEYWDYTAIFTRQDLC 322
Db 287 SLVVVKILYIEBQKGPVETVSNAAALTLRNFCAWKKLNKVSOKHPEYWDYTAIFTRQDLC 346
QY 323 GATTCDTGLMADYGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNHNDKVCVEEVEG 382
Db 347 GSHTCTDGLMADYGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNHNDKVCVEEVEG 406
QY 383 KLRANHMSFTLIQIDRANPWSACSAIITDFTDLSHGDCLLDOPSKPISLPELPGASY 442
Db 407 VTGDSHLMASLSSLDHSPWSPCSAYMVTSLDNGHGECCLMDKPNPILKPSDLPGLTY 466
QY 443 TLSQCELAFLGVGSKPCP--YMOYCTKLWCTGKAKGOMVCTRHFHWADGTSCEGKCLK 501
Db 467 DANROQFTFGEBSKPCDAASTCTTLWCTGSGVLLVCOTKHFHWADGTSCEGKCLK 526
QY 502 GACVERHNLNKH---RVDGSWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKGYCEGV 558
Db 527 GKCVNKTDM-KHFATPVHSGWGMWGPWGDSCRTCGGGVQYTWRECDNPVYKNGKGYCEG 585
QY 559 RVKRYSCNLEPCPSSASGSKFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCL 618
Db 586 RVKRYSCNLEPCPSSASGSKFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCL 644
QY 619 ICRANGTGYFYVLAPKVDGTLCSPDSTSVCVQGCICAKGDCNGLSKRFDKCGVCGGD 678
Db 645 TCEAKGIGYFYVLAPKVDGTLCSPDSTSVCVQGCICAKGDCNGLSKRFDKCGVCGGD 704
QY 679 NKSKKVTG 687
Db 705 GSTCKKMSG 713

RESULT 5
US-10-163-316-7
; Sequence 7, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Liebermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: MPI01-025P1RNW
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-163-316-7
Query Match 52.0%; Score 1947.5; DB 9; Length 968;
Best Local Similarity 52.3%; Pred. No. 3.6e-140;
Matches 381; Conservative 95; Mismatches 176; Indels 77; Gaps 16;
```

[illegible]

```

RESULT 6
US-09-965-631-6
; Sequence 6, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Fiddler, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucle
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 367
; TYPE: PRT
; ORGANISM: homo sapiens

```

US-09-965-631-6

Query Match 51.1%; Score 1913; DB 10; Length 367;
Best Local Similarity 99.7%; Pred. No. 4.7e-138;
Matches 363; Conservative 0; Mismatches 1; Indels 0

Qy	1	MLLIGILTLFAGRTAGGFEPEVVPVPIRLDPDINGRRYYWNGRPEDSGDGLIFOITAF	60
Db	1	MLLIGILTLFAGRTAGGFEPEVVPVIRLDPDINGRRYYWNGRPEDSGDGLIFOITAF	60
Qy	61	QEDFYHLTDPDAQFLAPAFSTEHLGVPLOGLTGSSDLRRCFFYSGDVNAEPDPSFAAVSLC	120
Db	61	QEDFYHLTDPDAQFLAPAFSTEHLGVPLOGLTGSSDLRRCFFYSGDVNAEPDPSFAAVSLC	120
Qy	121	GGLRGAFGYRGAEEYVISPLPNASAPAAQRNSQGAHLLQRRGVPGGSPGDPTSRCGVASGW	180
Db	121	GGLRGAFGYRGAEEYVISPLPNASAPAAQRNSQGAHLLQRRGVPGGSPGDPTSRCGVASGW	180
Qy	181	NPATILRALDYPKPRRAGFGESESRSSRRSGRAKRFVSPRYVETILWVADESMVKFHGADLEH	240
Db	181	NPATILRALDYPKPRRAGFGESESRSSRRSGRAKRFVSPRYVETILWVADESMVKFHGADLEH	240
Qy	241	YLLTLLATAARLYRHPISILNPINIVVVKVLLLRDSDGPKVTGNAALTLRNFCAWOKKLN	300
Db	241	YLLTLLATAARLYRHPISILNPINIVVVKVLLLRDSDGPKVTGNAALTLRNFCAWOKKLN	300
Qy	301	KVSDKHPEYWDTAILFTRQDLGGATTCDTLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTA	360
Db	301	KVSDKHPEYWDTAILFTRQDLGGATTCDTLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTA	360
Qy	361	HELG 364	
Db	361	HELG 364	

RESULT 7

US-10-09/-59/-1
: sequence 1 Application US/10097597

: Publication No. US20030022352A1

: GENERAL INFORMATION:

; APPLICANT: Hirose, Kunitaka

; APPLICANT: Inoguchi, Eiichi

; APPLICANT: HAKOZAKI, Michinori

; APPLICANT: Ishioka, Keiko

APPLICANT: MATSUSHITA KOU

APPLICANT: KUNO. KONTI.

;; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,

; TITLE OF INVENTION: pharmaceutical

; TITLE OF INVENTION: composition and

; FILE REFERENCE: Q57092

; CURRENT APPLICATION NUMBER: US/10/097,397
 ; CURRENT FILING DATE: 2002-03-15

; CURRENT FILING DATE: 2002-03-15
 ; PRIOR APPLICATION NUMBER: 09/445

: PRIOR FILING DATE: 1999-12-03

: PRIORITY APPLICATION NUMBER: JP 9-160422

; PRIOR FILING DATE: 1997-06-03

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PA

```

; SEQ ID NO 1
: LENGTH: 727

```

```

; LENGTH: 727
; TYPE: PRT

```

ORGANISM: Homo sapiens

US-10-097-597-1

Query Match	46.8%;	Score 1751;	DB.9;	Length 727;
Best Local Similarity	64.0%;	Pred. NO. 2.4e-125;		
Matches 309;	Conservative	68;	Mismatches 100;	Indels

QY 209 RAKRFVSIPRYVETLVWADESNWKFHGADLEHYLLTLLATAARLYKHPISILNPINIVVK 268
| | | | | | | | :
Db 9 RKKRFTSSPRYVETMLVAQDSMAEFHGSGLKHLLTLFESVAARLYKHPISRNSVSLVVK 68

QY 389 MMSPTLIQIDRANPWSACSAIIITDLDGSHGDCLLDQSPKIPSLPDLPGASYTLISQOC 448
 DB 189 MMSPTLIQIDRANPWSACSAIIITDLDGSHGDCLLDQSPKIPSLPDLPGASYTLISQOC 248
 QY 449 ELAFVGSKPCP-YMOYCTKLWCTGKAKGOMVQOTRHFPWADGTSCEGKLCCLKGCACVER 507
 DB 249 QFTFGEDSKHCPDAASTCTLWCTGSGGLVVCOTKHFPWADGTSCEGKLCCLKGCACVER 308
 QY 508 HNLNKH---RVDGSAKWDPYGPCSRTCCGGVQGLARROCTNPTPANGKYGCEGVRYKRS 564
 DB 309 TD-RKHEDTFFHGSWGPWGPCSRTCCGGVQGLARROCTNPTPANGKYGCEGVRYKRS 367
 QY 565 CNLEPCSSASGSKFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANG 624
 DB 624 CNLEPCSSASGSKFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANG 624
 QY 625 TGYFVVLAPKVVDTGLSPDSTSVCVQGCICAKGDCNGLSKRFRDKCGVCGGDNKSCCK 684
 DB 625 TGYFVVLAPKVVDTGLSPDSTSVCVQGCICAKGDCNGLSKRFRDKCGVCGGDNKSCCK 684
 QY 685 VTG 687
 DB 487 ISG 489

RESULT 10
 US-10-097-597-12
 ; Sequence 12, Application US/10097597
 ; Publication No. US20030022352A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirose, Kunitaka
 ; APPLICANT: Inoguchi, Eiichi
 ; APPLICANT: Hakozaaki, Michinori
 ; APPLICANT: Ishioka, Keiko
 ; APPLICANT: Matsushima, Kouji
 ; APPLICANT: Kuno, Kouji
 ; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
 ; pharmaceutical
 ; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
 ; FILE REFERENCE: Q57092
 ; CURRENT APPLICATION NUMBER: US/10/097,597
 ; PRIOR FILING DATE: 2002-03-15
 ; PRIOR APPLICATION NUMBER: 09/445,023
 ; PRIOR FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 727
 ; TYPE: PRT
 ; ORGANISM: Mus sp.
 US-10-097-597-12

Query Match 46.4%; Score 1738; DB 9; Length 727;
 Best Local Similarity 63.4%; Pred. No. 2.3e-124;
 Matches 306; Conservative 68; Mismatches 103; Indels 6; Gaps 4;

QY 209 RAKRFVSPRYVETLVVADESMVKFHGADLEHYLLTLATAARLYRHPISILNPIVYVK 268
 DB 9 RKKRFVSPRYVETLVVADESMVKFHGADLEHYLLTLATAARLYRHPISILNPIVYVK 68

QY 269 VLLLRDRSGPKVTGNAALTNRNFCQWKLKNSDKHPEYWDTAIFTRQDLGGATCD 328
 DB 69 ILVIYEEQKGPVTSNAALTNRNFCQWKLKNSDKHPEYWDTAIFTRQDLGGATCD 128

QY 329 TLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNVKVEEYFGLRANH 388
 DB 129 TLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNVKVEEYFGLRANH 388

QY 389 MMSPTLIQIDRANPWSACSAIIITDLDGSHGDCLLDQSPKIPSLPDLPGASYTLISQOC 448
 DB 448 MMSPTLIQIDRANPWSACSAIIITDLDGSHGDCLLDQSPKIPSLPDLPGASYTLISQOC 448

DB 189 LMASMLSSLDHSPWSPCSAYMWTSFLDNGHGECLMDKPQNPIKLPSDLPTGLYDANRQC 248
 QY 449 ELAFVGSKPCP-YMOYCTKLWCTGKAKGOMVQOTRHFPWADGTSCEGKLCCLKGCACVER 507
 DB 249 QFTFGEDSKHCPDAASTCTLWCTGSGGLVVCOTKHFPWADGTSCEGKLCCLKGCACVER 308
 QY 508 HNLNKH---RVDGSAKWDPYGPCSRTCCGGVQGLARROCTNPTPANGKYGCEGVRYKRS 564
 DB 309 TD-RKHEDTFFHGSWGPWGPCSRTCCGGVQGLARROCTNPTPANGKYGCEGVRYKRS 367
 QY 565 CNLEPCSSASGSKFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANG 624
 DB 624 CNLEPCSSASGSKFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANG 624
 QY 625 TGYFVVLAPKVVDTGLSPDSTSVCVQGCICAKGDCNGLSKRFRDKCGVCGGDNKSCCK 684
 DB 625 TGYFVVLAPKVVDTGLSPDSTSVCVQGCICAKGDCNGLSKRFRDKCGVCGGDNKSCCK 684
 QY 685 VTG 687
 DB 487 MSG 489

RESULT 11
 US-10-097-580-12
 ; Sequence 12, Application US/10097580
 ; Publication No. US20030032168A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirose, Kunitaka
 ; APPLICANT: Inoguchi, Eiichi
 ; APPLICANT: Hakozaaki, Michinori
 ; APPLICANT: Ishioka, Keiko
 ; APPLICANT: Matsushima, Kouji
 ; APPLICANT: Kuno, Kouji
 ; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
 ; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
 ; FILE REFERENCE: Q57092
 ; CURRENT APPLICATION NUMBER: US/10/097,580
 ; PRIOR FILING DATE: 2002-03-15
 ; PRIOR APPLICATION NUMBER: 09/445,023
 ; PRIOR FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 727
 ; TYPE: PRT
 ; ORGANISM: Mus sp.
 US-10-097-580-12

Query Match 46.4%; Score 1738; DB 9; Length 727;
 Best Local Similarity 63.4%; Pred. No. 2.3e-124;
 Matches 306; Conservative 68; Mismatches 103; Indels 6; Gaps 4;

QY 209 RAKRFVSPRYVETLVVADESMVKFHGADLEHYLLTLATAARLYRHPISILNPIVYVK 268
 DB 9 RKKRFVSPRYVETLVVADESMVKFHGADLEHYLLTLATAARLYRHPISILNPIVYVK 68

QY 269 VLLLRDRSGPKVTGNAALTNRNFCQWKLKNSDKHPEYWDTAIFTRQDLGGATCD 328
 DB 69 ILVIYEEQKGPVTSNAALTNRNFCQWKLKNSDKHPEYWDTAIFTRQDLGGATCD 128

QY 329 TLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNVKVEEYFGLRANH 388
 DB 129 TLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNVKVEEYFGLRANH 388

QY 389 MMSPTLIQIDRANPWSACSAIIITDLDGSHGDCLLDQSPKIPSLPDLPGASYTLISQOC 448
 DB 448 MMSPTLIQIDRANPWSACSAIIITDLDGSHGDCLLDQSPKIPSLPDLPGASYTLISQOC 448

QY 449 ELAFVGSKPCP-YMOYCTKLWCTGKAKGOMVQOTRHFPWADGTSCEGKLCCLKGCACVER 507

```
Db 249 QFTFGESKHCDAASTCTTTLWCTSGGLLVCTQKHFPMADGTSCEGKWCYSGKCVNK 308
QY 508 HNLNKH---RVDSWAKWDPYGPCSRTCGGGVQLARRQCTNPTTPANGKCYCEGVKRYKRS 564
Db 309 TDM-KHFATPVHSGWPGMGDCSRTCGGGVQVYTMRECDNPVKNGKCYCEGKRVYRS 367
QY 565 CNLEPCSSASGKSPREOCEAFNGYNHSTNRLTLAVAVPKYSGVSPDRCKLICRANG 624
Db 368 CNIEDCPDN-NGKTFREOCEAHNEFSKASFGNEPTVETWPKYAGVSPDRCKLICRANG 426
QY 625 TGFFVYLPKVVVDGTLCSPDSTSVCVQGCIKAGCDGNLGSKKRFDKCGVCGGDNKSKCK 684
Db 427 IGYFFVLPKVVVDGTPCSPDSTSVCVQGCYKAGCDRIIDSKKFKDKCGVCGGNGSTCKK 486
QY 685 VTG 687
Db 487 MSG 489

RESULT 12
US-09-445-023A-12
; Sequence 12, Application US/09445023A
; Patent No. US2002019167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/09/445,023A
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-445-023A-12

Query Match 46.4%; Score 1738; DB 10; Length 727;
Best Local Similarity 63.4%; Pred. No. 2.3e-124;
Matches 306; Conservative 68; Mismatches 103; Indels 6; Gaps 4;

QY 209 RAKRFVSPRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPSTILNPINIVVK 268
Db 9 RKRFVSPRYVETLVVADQSMADFHGSLGHLTLTLFVAARLYRHPSTIRNSISLVVK 68
QY 269 VLLLRDRSGPKVTGNAAALTLRNFCWAKQKLNKYSKHPYWDTAIFTRQDLCSGATCD 328
Db 69 ILVIEYKQGEVTSNAALTLRNFCWAKQKLNKYSKHPYWDTAIFTRQDLCSGATCD 128
QY 329 TLGMADVGTCDPKRSCSVIEDDGLPSAFTAHGLHVFNNPNDHNVKVCSEVFGKLRANH 388
Db 129 TLGMADVGTCDPKRSCSVIEDDGLQAAFTAHGLHVFNNPNDHNVKVCSEVFGKLRANH 188
QY 389 MNSPTLIQIDRANPWSACSAIIFDLDGSHGDCLLDPSKPISLPDLPGASYTLSSQOC 448
Db 189 LWSKMLSSLDHSDQWSPSCAYMTSFLDNGHGECLMDKPNQPKLSDLPGLTYLDNRQC 248
QY 449 ELAFGVGSKPCP-YMQVCTKLWCTGKAKGMVQOTRHFHPWADGTSCEGKLCCLKGACVER 507
Db 249 QFTFGESKHCDAASTCTTTLWCTSGGLLVCTQKHFPMADGTSCEGKWCYSGKCVNK 308
QY 508 HNLNKH---RVDSWAKWDPYGPCSRTCGGGVQLARRQCTNPTTPANGKCYCEGVKRYKRS 564
```

```
Db 309 TDM-KHFATPVHSGWPGMGDCSRTCGGGVQVYTMRECDNPVKNGKCYCEGKRVYRS 367
QY 565 CNLEPCSSASGKSPREOCEAFNGYNHSTNRLTLAVAVPKYSGVSPDRCKLICRANG 624
Db 368 CNIEDCPDN-NGKTFREOCEAHNEFSKASFGNEPTVETWPKYAGVSPDRCKLICRANG 426
QY 625 TGFFVYLPKVVVDGTLCSPDSTSVCVQGCIKAGCDGNLGSKKRFDKCGVCGGDNKSKCK 684
Db 427 IGYFFVLPKVVVDGTPCSPDSTSVCVQGCYKAGCDRIIDSKKFKDKCGVCGGNGSTCKK 486
QY 685 VTG 687
Db 487 MSG 489

RESULT 13
US-10-174-590-352
; Sequence 352, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P343031C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-352

Query Match 45.1%; Score 1688.5; DB 9; Length 837;
Best Local Similarity 47.2%; Pred. No. 1.6e-120;
Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;

QY 1 MLLLTGLTLAFAGRTAGGPEPEREVVPIRLDPDINGRYYWRGPEDSGDLGFIQTAF 60
Db 37 LLLLLASLLPSARLASPLPREEEIVFPEKLGSLV-----PGSGAPARLLCRLQAF 88
QY 61 QEDFYLHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDSFAAVSLC 120
Db 89 GETLLELEQDSGVQVEGLTVQVIGQAPL-LGGAEP--GTVLTGTINGDPESVSLHWD 145
QY 121 GG-LRCAGFYGRAEYVISPLPNASAPAAORNQGAHLQRRGVPGGSDPSTRSGVAG 179
Db 146 GGALLGVLOYRAELHLQLEGGTPNSA--GGGAHLRKK---SPASGGQPMCNV--- 196
QY 180 WNPAILRALDPYKPRRAGFESRRSRGRKRFRYSIPRYVETLVVADESVMKFGADLE 239
Db 197 -----KAPLGSPSPRPR--RAKRFASLSRFVETLVVADDKMAAFHAGLJK 239
QY 240 HYLLTLTAARLYRHPSTILNPINIVVKVLLLRDRSGPKVTGNAAALTLRNFCWAKQK 299
Db 240 RYLLTYMAAAKAFKHPISIRNPVSLVTVTLVILGSGEGPQVGPAAQTLRSFCWAGRL 299
QY 300 NKVSKHPEYWDTAIFTRQDLCSGATCDTLGMADVGTCDPKRSCSVIEDDGLPSAFT 359
Db 300 NTPEDSGPDHFDTAIFTRQDLCSGATCDTLGMADVGTCDPARSCAIVEDDGLSAFTA 359
QY 360 AHELGHVFNNPNDHNVKVCSEVFGKL-RANHMSPTLIQIDRANPWSACSAIITDLDG 418
```

Db 360 AHELGHVENMLHNSKPCISLNGPLSTRHVMAVMAHVDPEEPWSPCSARFITDFLNG 419
QY 419 HGDCLLDQSPKPSISLPEDLPQASYTLISOCELAFGVSKPCPYM-QYCTKLWCTGKAKQ 477
Db 420 YGHCLLDKPEAPLHLPVTFPGDYADROCLTFGPDSSRHRCPLPPPCAALWCSGHLNGH 479
QY 478 MVOCTRHFPWADGTCGEGKCLKLGACVERHNLNKHVD--GSAKWDPYGPCSRTCCGG 535
Db 480 AMQCTKHSFWADGTCGPAQACMGRCGLHMDQLQDNIPQAGGWGPMGDCSRTCCGG 539
QY 536 VOLARRQCTNPTPANGKYCEGVVRYRSCNLEPCPSASGSKSFREOCEAFNGYHNSTN 595
Db 540 VQFSRSDCTRPVPRNGKYCEGRRTRFRSCNTEDCP-TGSALTFRBOCAA---YHNRTD 595
QY 596 RLTL--AVAWPKYSGVSPRDKCLKICRANGTGYFYVLAPKVVDGTLCSPDSTSVCVQG 652
Db 596 LKSPFGPMWVRYTGVAPDOCKLTQCARALGYVYVLEPRVVDGTPCSPDSSSVCVQG 655
QY 653 KCICAGDGNLGSKKRFDKCGVCGDNKSKKVTG 687
Db 656 RCIHAGCDRIIGSKKFKDKCMVCGDGGCGSKQSG 690

RESULT 14

US-10-176-758-352
; Sequence 352, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176, 758
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-352

Query Match 45.1%; Score 1688.5; DB 9; Length 837;
Best Local Similarity 47.2%; Pred. No. 1.6e-120;
Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRYYWRGPDSDGQGLIFQTAF 60
Db 37 LLLLLLASLPSARLASPLPREEEIVPEKLSVLT-----PGSGAPARLLCRLQAF 88
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRCFCYSGDYNAPDSFAAVSLC 120
Db 89 GETLLELEQDSGVQVEGLTVQYLGAPE-LLGGAEP--GTYLGTGTNGDPESVASLHWD 145
QY 121 GG-LRGAFGYGAEVTSIPLENASAPAAORNSQGAHLQRRGVPGSPGDPSTSRCSVASG 179
Db 146 GGALLGVQYGAELHLQPLEGGTPNSA--CGPAHLIRK-----SPASGGPMCNV--- 196
QY 180 WNPAILRALDPKPRAGFGESRRSRAKRFVSIPIRYVETILVVADESMKVFHGDLE 239
Db 197 -----KAPLGSPSPRPR--RAKREASLSRFVETILVWADKMAAFHAGLK 239

QY 240 HYLLTLATARLYRHSILNPINIVVVKVLLLRDRDSGPKYTGNAALTLRNECAQKKL 299
Db 240 RYLLTVMAAAKAKFHSIRNPVSLVTRVLILGSGEGPOVGPSSAAQTLRSCAMORGL 299
QY 300 NKVSKHPEYWDTAITLFTRODLGATTCDTLGMADVGTMDCKPRKSCSVIEDDGLPSAFTT 359
Db 300 NTPEDSGPDHFDTAITLFTRODLGATTCDTLGMADVGTMDCKPRKSCSVIEDDGLPSAFTT 359
QY 360 AHELGHVENMHPDNKVCEEVFGKL-RANHMSPTLQIDRANPWSACSAAITDFLDSG 418
Db 360 AHELGHVENMHLNKSCKPCISLNGPLSTRHVMAVMAHVDPEEPWSPCSARFITDFLNG 419
QY 419 HGDCLLDQSPKPSISLPEDLPQASYTLISOCELAFGVSKPCPYM-QYCTKLWCTGKAKQ 477
Db 420 YGHCLLDKPEAPLHLPVTFPGDYADROCLTFGPDSSRHRCPLPPPCAALWCSGHLNGH 479
QY 478 MVOCTRHFPWADGTCGEGKCLKLGACVERHNLNKHVD--GSAKWDPYGPCSRTCCGG 535
Db 480 AMQCTKHSFWADGTCGPAQACMGRCGLHMDQLQDNIPQAGGWGPMGDCSRTCCGG 539
QY 536 VOLARRQCTNPTPANGKYCEGVVRYRSCNLEPCPSASGSKSFREOCEAFNGYHNSTN 595
Db 540 VQFSRSDCTRPVPRNGKYCEGRRTRFRSCNTEDCP-TGSALTFRBOCAA---YHNRTD 595
QY 596 RLTL--AVAWPKYSGVSPRDKCLKICRANGTGYFYVLAPKVVDGTLCSPDSTSVCVQG 652
Db 596 LKSPFGPMWVRYTGVAPDOCKLTQCARALGYVYVLEPRVVDGTPCSPDSSSVCVQG 655
QY 653 KCICAGDGNLGSKKRFDKCGVCGDNKSKKVTG 687
Db 656 RCIHAGCDRIIGSKKFKDKCMVCGDGGCGSKQSG 690

RESULT 15

US-10-175-737-352
; Sequence 352, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-352

Query Match 45.1%; Score 1688.5; DB 9; Length 837;
Best Local Similarity 47.2%; Pred. No. 1.6e-120;
Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRYYWRGPDSDGQGLIFQTAF 60
Db 37 LLLLLLASLPSARLASPLPREEEIVPEKLSVLT-----PGSGAPARLLCRLQAF 88
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRCFCYSGDYNAPDSFAAVSLC 120
Db 89 GETLLELEQDSGVQVEGLTVQYLGAPE-LLGGAEP--GTYLGTGTNGDPESVASLHWD 145

QY 121 GG-LRGAFYRGAEYVISPLPNASAPAAQNSOGAHLLORRGYPGGPSGDPTRSGYASG 179
Db 146 GGALLGVLYRGAEHLQLPLEGGTPNSA--GGFGAHLRRK-----SPASGGQPMCNV--- 196
QY 180 WNPAILRALDPYKPRRAGESRRRRSGRAKRFVSIPIRYVETLVVADESVMKFHGADLE 239
Db 197 -----KAPLGSPSPRPR--RAKRFASLSRFVETLVVADDDKMAAFHGAGLK 239
QY 240 HYLLTLATAARLYRHPSILNINIVVKKVLLLRDRDSGPKVTGNAALTLRNFCAWQKKL 299
Db 240 RYLLTYMAAAAKAFKHPISRNVPVSLVTRVLGSGEGEPQVGPSSAAQTLRSFCAWQRL 299
QY 300 NKVSDRHPYWDTAILFTQDLCGATTCTDLGMADVMTCDPKRSCSVIEDDGLPSAFTT 359
Db 300 NTFEDSGPHFTAILFTQDLCGVSTCDTLGMADVMTVCDPARCAIVEDDGLQSAFTA 359
QY 360 AHELGHVFNHNDNVKVEEVEFGKL-RANHMSPTLIQIDRANPWSACSAAIITDFLDG 418
Db 360 AHELGHVFNHNDNSKPCISLNGPLTSRHHVMAVYMAHYDPEEPWSPCSAREITDFLDNG 419
QY 419 HGDCILLDQSPKPLSLPEDLPGASYTLSQOCELAFGVGSKPCPYM-OYCTKLMCTGAKGQ 477
Db 420 YGCHLLDKPEAPLHLVTFPPGKDYDADROQLTFGPDSTRHCPQLPPPCAAALWCSGHLNGH 479
QY 478 MVCOTRHPFWADGTCGEGKCLKLGACVERHNLNKHRYD--GSAKWDPYGPSCSRTCGG 535
Db 480 AMCOTKHSFWADGTCGPAQACWGGRCCLHMDQLQDFNIPQAGGFWGFWGDCSRTCGG 539
QY 536 VQLARQCTNPTPANGKYCEGVRYKYSCLNEPCPSSASGKSFREEQCEAFNGYNHSTN 595
Db 540 VQFSRDCTRPVPRNGKYCEGRRTRFRSCNTEDCP-TGSALTFREEQCAA---YNHRTD 595
QY 596 RLTL---AVAWPKYSGVSPRKCLICRANGTGYFYVLAPKVVDGTLCSDPDSTSVCVQG 652
Db 596 LFKSFFGPMDWVPRYTGVAPOQCKLTQCARALGYYYVLEPRVVDGTPCSPDSSSVCVQG 655
QY 653 KCIKAGCDGNLGSKKRFDKCGVCGDNKSKKVTG 687
Db 656 RCIHAGCDRIIGSKKKFKCMVCGDGCCKSQSG 690

Search completed: April 29, 2003, 17:21:58
Job time : 20.2713 secs

GenCore version 5.1.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 15.8904 Seconds
(without alignments)
4156.253 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_687

Perfect score: 3744

Sequence: 1 MLLLGILTLAFAGRTAGGFE.....RFDKCGVCGDNKSKKVTG 687

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1947.5	52.0	951	2 T00017	gene ADAMTS-1 prot
2	1697.5	45.1	837	2 T00355	hypothetical prote
3	1223	32.7	2165	2 T21371	hypothetical prote
4	1084	29.0	550	2 T47158	hypothetical prote
5	980.5	26.2	1205	2 T18517	procollagen N-endo
6	479	12.8	1444	2 T18856	hypothetical prote
7	426.5	11.4	860	2 T16892	hypothetical prote
8	406.5	10.9	1558	2 C89114	protein C37C3.6a l
9	406.5	10.9	2167	2 T34395	hypothetical prote
10	362	9.7	957	2 T15976	hypothetical prote
11	344.5	9.2	571	2 S24789	hypothetical prote
12	337	9.0	609	2 S55270	catrocollastatin p
13	329	8.8	903	2 S60257	metrin alpha - mo
14	323.5	8.6	617	2 S48160	metalloproteinase
15	313	8.4	549	2 S48169	metalloproteinase
16	313	8.4	826	2 A60385	monocyte surface a
17	309.5	8.3	951	2 T00260	hypothetical prote
18	307.5	8.2	789	2 S28259	androgen-regulated
19	307	8.2	411	1 HYSNFA	fibronectin
20	304	8.1	407	2 S66260	metalloproteinase
21	300.5	8.0	616	2 A55796	ecarin precursor -
22	297	7.9	480	1 A30065	trigamin precursor
23	292	7.8	478	2 JC4880	fibronolytic metal
24	290.5	7.8	481	2 JC4342	fibronolytic prote
25	289	7.7	414	2 S41609	atrolysin C (EC 3.
26	286.5	7.7	610	2 JC7530	vascular apoptosis
27	284	7.6	414	1 HYSAC	atrolysin C (EC 3.
28	280	7.5	481	2 S43125	trimucin precursor
29	277.5	7.4	814	2 G02390	disintegrin-like m

30	273.5	7.3	478	2 A43296	atrolysin E (EC 3.
31	265	7.1	414	2 S41608	atrolysin B (EC 3.
32	258.5	6.9	670	2 I65967	disintegrin-like m
33	257	6.9	776	2 S28258	androgen-regulated
34	251	6.7	1042	2 T26644	hypothetical prote
35	248.5	6.6	1059	2 T22545	hypothetical prote
36	248.5	6.6	1170	1 TSHUP1	thrombospondin 1 p
37	245.5	6.6	1170	2 A40558	thrombospondin 1 p
38	244.5	6.5	419	2 A59414	metalloproteinase
39	244	6.5	952	2 T18900	disintegrin and me
40	243	6.5	478	2 JQ1301	hemorrhagic protei
41	234.5	6.3	508	2 T22836	hypothetical prote
42	233.5	6.2	419	2 S41607	atrolysin A (EC 3.
43	232	6.2	823	2 S18968	cyritestin precurs
44	228	6.1	735	2 G02937	fertilin beta - cr
45	220	5.9	416	2 A37877	hemorrhagic protei

ALIGNMENTS

RESULT 1

T00017

gene ADAMTS-1 protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00017

R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.

Genomics 46, 466-471, 1997

A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 g

A:Reference number: Z14055; MUID:98110583; PMID:9441751

A:Accession: T00017

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-951 <KUN>

A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057

A:Superfamily: thrombospondin type 1 repeat homology

C:Experimental source: strain 129SVJ

C:Genetics:

A:Gene: ADAMTS-1

A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2

F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match

Best Local Similarity 52.0%; Score 1947.5; DB 2; Length 951;

Matches 381; Conservative 95; Mismatches 176; Indels 77; Gaps 16;

QY 1 MLLLGILTLAFAGRTAGG--FEPEVYVPIRLDPDINGRRYWRGP-EDSGDQGLIFOI 57

Db 20 LLLASITMLLCARGAHRPTDEELVLP-SLE-----RAPHGDSITTRL--RL 66

QY 58 TAFQEDFYHLTPDQAQFLAPAFSTHGLVPLQGLTGG-----SDLRRCFYSGDVNA 109

Db 67 DAFQOOLHLKLPDQSGFLAPGFTLTQV-----GRSPGEAQLHLDPTGLAHCFYSGTVNG 121

QY 110 EPDFAAYSLCGLRGAFYGAEYVISPPLNAS-----APAAQRNSQGA----HLQRRG 161

Db 122 DFGSAALSICGVGAFYLGQEEFFIQAPGVATERLAPVPEESSARFOFHILRRR- 180

QY 162 VPGSGSGDPTSCGVASGNPAILRALDPYKPRFAGFGESESRRR----- 206

Db 181 ----RRSGGAKCGYMD-----DETLPSTDSRPSQNTNRQWVPRDPTQDACKP 226

QY 207 ---SGRAKRFVSIPIRYETLWVADESMVKFHGADLEHYLLTLLTAARLYRHPSILNPI 262

Db 227 SGPGSIRKKRFVSSPRYETMLVADQSMADFHGSLKHYLLTLFSVAARFYKHPISIRNSI 286

QY 263 NIVVYKVLRLDRDGPVKVTGNAALTNRFCWAKKLNKSKDKHPEYWDATILTRDLC 322

Db 287 SILVVYKILVIEQKGPVTSNAALTNRFCWAKKLNKSKDKHPEYWDATILTRDLC 346

QY 323 GATTCDDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFG 382

Db 323 GATTCDDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFG 382

Db 347 GSHTCDTLGMADVTGCDPSRSCSVIEDDGLQAFAAHTAHELGHVFNMPHDDAKHCASLNG 406
QY 383 KLRANHMMSPTLLQIDRANPWSACSAIITDFLDGSHGDCILDDQSPKPTSLPDLPGASY 442
Db 407 VTGDSHLMASLSSLDHSPQSPSCSYWTSFLDNGHGCLMDKPKONPKLPSLDPLGTL 466
QY 443 TLSQOCELAFAGSKPCP-YMOTCTKLWCTGKAGQWVQOTRHFPWADGTCGEGKLCUK 501
Db 467 DANROCOFTGESKHCPOAASCTTLWCTGTSGGLLVQCTKHPWADGTCGEGKWCVS 526
QY 502 GACVERHNLKHX--RVDSNAKWDPYGPCSRTCTCGGVQVLARRQCTNPTPANGKYCEGV 558
Db 527 GKCVNKTDM-KHEATPVHGSWGPWGDCSRTCTCGGVQVYTWRECDNPPKNGKYCEGK 585
QY 559 RVKVRSCNLEPCPSASGSKFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKL 618
Db 586 RVRYSRNCIEDCDPN-NGKTFREEQCEAHNEFSKAFSGNEPTVETWPKYAGVSPDKCKL 644
QY 619 ICRANGTGYFVLAPKVVDTGLSPDSTSVCGQKCIKAGCDGNLGSKKRFDKCGVCGD 678
Db 645 TCEAKGIGYFVLQPKVVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKFDKCGVCGG 704
QY 679 NKSCKKVTG 687
Db 705 GSTCKRMSG 713

RESULT 2
T00355
hypothetical protein KIAA0688 - human
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00355
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: Z14142; MUID:98403880; PMID:9734811
A:Accession: T00355
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-837 <ISH>
A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BA031663.1; PID:g3327190
A:Experimental source: Brain
C:Genetics:
A:Gene: KIAA0688
A:Superfamily: thrombospondin type 1 repeat homology
F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 45.18; Score 1687.5; DB 2; Length 837;
Best Local Similarity 47.28; Pred. No. 5.5e-112; Mismatches 215; Indels 49; Gaps 14;
Matches 328; Conservative 103

QY 1 MLLGILTLFAGRTAGTGFEPREVVPIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
Db 37 LLLLLLASLLSARLASPLPREIEIVPEKLSVLF-----PGSGTTPARLLCRLOAF 88
QY 61 QEDFYVLTPOAFLAPAFSTEHLGVPLQGLTGGSSDLRCFCYSGDYNAPDSPAAYSLC 120
Db 89 GETLLELEQDSGVQVEGLTVQYLQAPE-LLGGAEP--GTYLGTNGDPESVASHWD 145
QY 121 GG-LRGAFGYGAEVVISPLPNASAPAAORNSQGAHLORRGVPGSPGDPTRSCGVASG 179
Db 146 GALLGVLOYGAELHLOPLGGTGNPSA--CGPAHILRRK-----SPASGGPMCNV--- 196
QY 180 WNPAILRALDYPKPRRAGGESRRRRSGRAKRFVSPRYVETVLVVADESVMKPHGADLE 239
Db 197 -----KAPLGSPPRPR--RAKREASLSRFVETVLVVADEKMAAFHAGLAK 239
QY 240 HYLLTLLAARLYRHPISILNPINIVVYKVLRLDRDSGPKVTGNAALTLENFCAWOKKL 299
Db 240 RYLLTVMAAAKAFKHPISIRNPVSLVTRVLILSGEGEGPQVGPSPAQTLRSFCAWOGL 299
QY 300 NKVSDKHPEYWDITLFTRODLCGATCTDITLGMADVTGTCMDPKRSCSVIEDDGLPSAFT 359

Db 300 NTPEDSDPDHEDTALLFRQDLQGVSTCDILGMADVTGCDPARSCAIVEDDGLQSAFTA 359
QY 360 AHELGHVENPHDNVYCEEVEFGKL-RANHMMSPTLLQIDRANPWSACSAIITDFLDG 418
Db 360 AHELGHVENHMLHDSKPCISLNGPLSTSRHVMAPVMAHVDEPEWSPCSARFIFDLDNG 419
QY 419 HGDCILLDOPSPISLPEDLPQASTYLSQOCELAFAGVSKPCPYM-QYCTKLWCTGKAKQ 477
Db 420 YGHCILLDRPEAPLHLPVTFPGKYDADQCCQLTFGPSRHCPLPPPCAALWMSGHLNGH 479
QY 478 MVQOTRHPHWDGTCGEGKLCAGACVERHNLKHRVD--GSNAKWDPYGPCSRTCTCGG 535
Db 480 AMCQTKHSPWADGTPCGPAQACMGRCRLHMDQLQDFNIPQAGGHPGHPWGDCCSRTCTCGG 539
QY 536 VOLARRQCTNPTPANGKYCEGVRYKYSRNCLEPCPSASGSKFREEOCEAFNGYNHSTN 595
Db 540 VQFSRSDCTRPVPRNGKYCEGRRTRFRSCNTEDCP-TGSALTFRFEEQCAA---YNHRTD 595
QY 596 RLTL---AVAWPKYSGVSPRDKCKLICRANGTGYFYVLAPKVVDTGLCSPDSTSVCG 652
Db 596 LFKSPFGMDWVPRYTGVAPODQCKLTQARALGYYYVLEPRVVDGTPCSPDSSSVCG 655
QY 653 KCIKAGCDGNLGSKKRFDKCGVCGGDNKSKKVTG 687
Db 656 RCIRHAGCDRIIGSKKKFKDKCVCGGDSGCKSKQG 690

RESULT 3
T21371
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21371; T24896
R:Galadsty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone F25H8
R:Galadsty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <W12>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
C:Genetics:
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/2; 634/2; 744/1;

Query Match 32.7%; Score 1223; DB 2; Length 2165;
Best Local Similarity 40.5%; Pred. No. 1.7e-78;
Matches 257; Conservative 87; Mismatches 214; Indels 76; Gaps 16;

QY 101 CFYSGDVNAEPDSPAFAVSLC---GGLRGAFGYGAEVVISPLPNASA---PAAORNSQA 154
Db 169 CIYRAHVKG-VHQHSIVNLCDSEDGLYGLMALPSGIHTVEITISNGTEHDGASRHRQ-- 225
QY 155 HLQORGVCGPGSDPTSCGV-----ASGNPAILRALDYPKPRRAGGESRSR 204
Db 226 HLVRKFDPMHFKSFDHLNSTSVNETETTVATWQDQEDVIER-----KARSR 272
QY 205 RRSGRKRRFYSIPRYVETVLVVADESVMKPHGADLEHYLLTLLAARLYRHPISILNPIN 264
Db 273 RAAN-----SWDHYVEVLVVAOTKMVEYHGRSLDEYVLTFLFTVASIYRHQSLASIN 326

Db 382 ---KSDLTFTGFCGSKWQGLRCVPMWGTNEIQPTVQHVAPVVTLPSPRIDGWSG 438
QY 421 DCLDQ-----PSKIPSLPDLPGAS-----YTLSSOCELA----- 451
Db 439 ATICSQTCNGILGVLGATARTCSAPYANGSGDCVSTRAVLCRSQCGRASKSVDE 498
QY 452 -----FVGSKPCPYMYQYCKLWCTGKAKGOMVCQTRHFPW-----A 488
Db 499 YISDKMEQKRLKNDRELTGKGLNRFQPCACKVFCVDQ---QHYGSRNYRFFGDNLP 555
QY 489 DGTSCGEGKCLLKGACVERINLKHRYD-----GSAKWADPYGP 527
Db 556 DGTSCGVDYCLDGEALCNMNLALSRDOSCPDTCTPTDSSSVYRGOWTSLWTS 615
QY 528 CSRTCGGCVQLARRQCTNPTPANGGKYCEGVVRKYRSCNLEPCP----- 571
Db 616 CTATCGGGYKRNKACS-----ITG--QCEGNEDETEVCSSSESCPSVLRVGNKSWTWN 669
QY 572 -----SSASGKSFREQC-----EAFN--CYNHSTNRLTLAVAWPKYSGVSPRDKCLI 619
Db 670 HCSVSCGRGQARYRKCLSPHRLAFCGKNTKVTNELRTFFKARSYIMCSVR--CNKI 727
QY 620 CR-----ANGTGYFVYVLPAPKVVVDGTLCSPTDSTSVCVGKCIKAGCDGN 662
Db 728 KRNTISEKNIEVRSCDNGPCNAIGVGTGCGWSTCSTSCGPG--TLVRQTCNREPCDGS 785
QY 663 LGSKKRFDKCV---CGDN-----KSKKYTG 687
Db 786 AHERR---SCNVATCQNDGIWLSLWNSDCSRVCG 817

RESULT 11

S24789
Jararagin C precursor - Jararaca (fragment)
N:Alternate names: single chain botrocetin
N:Contains: disintegrin-like 28k protein; hemorrhagic proteinase (EC 3.4.24.-)
C:Species: Bothrops jararaca (Jararaca)
C>Date: 20-Feb-1995 #sequence_revision 29-Aug-1997 #text_change 09-Jun-2000
C:Accession: S24789; J02245; A4463; A37958; J02373
R:Palme, M.J.I.
submitted to the EMBL data library, August 1992
A:Reference number: S24789
A:Accession: S24789
A:Molecule type: mRNA
A:Residues: 1-571 <PAI>
A:Cross-references: EMBL:X68251; NID:g62467; PID:g62468
R:Usami, Y.; Fujimura, Y.; Miura, S.; Shima, H.; Yoshida, E.; Yoshioka, A.; Hirano, K.;
Biochem. Biophys. Res. Commun. 201, 331-339, 1994
A:Title: A 28 kDa-protein with disintegrin-like structure (Jararagin-C) purified from B
A:Reference number: J02245; MUID:94256999; PMID:8198592
A:Accession: J02245
A:Molecule type: protein
A:Residues: 360-571 <USA>
A:Experimental source: venom
R:Palme, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.
J. Biol. Chem. 267, 22869-22876, 1992
A:Title: Purification, cloning, and molecular characterization of a high molecular weight
illy.
A:Reference number: A4463; MUID:93054601; PMID:1385408
A:Accession: A4463
A:Molecule type: mRNA
A:Residues: 1-23, 'Q', 25-92, 'G', 94-131, 'G', 133-169, 'Q', 171-571 <PA2>
A:Cross-references: GB:X68251
A:Experimental source: venom gland
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIP:118104)
R:Fujimura, Y.; Titani, K.; Usami, Y.; Suzuki, M.; Oyama, R.; Matsui, T.; Fukui, H.; Sug
Biochemistry 30, 1957-1964, 1991
A:Title: Isolation and chemical characterization of two structurally and functionally di
A:Reference number: A37958; MUID:91129280; PMID:1993206
A:Accession: A37958
A:Molecule type: protein
A:Residues: 360-372, 'E', 374-378, 'X', 380-384 <FUJ>

A:Note: 361-Val was also found
C:Comment: Inhibits collagen- and ADP-induced platelet aggregation.
C:Superfamily: mouse meltrin alpha; disintegrin homology
C:Keywords: hydrolase; metalloproteinase; venom; zinc
F:360-571/Product: Jararagin C #status experimental <MAT>
F:362-444/Domains: disintegrin homology <DIS>
F:295,299,305/Binding site: zinc (His) #status predicted
F:296/Active site: Glu #status predicted

Query Match 9.2%; Score 344.5; DB 2; Length 571;
Best Local Similarity 22.7%; Pred. No. 9.3e-17;
Matches 160; Conservative 88; Mismatches 256; Indels 201; Gaps 35;
QY 12 AGRTAGGPEPEREVVPIRLDPDINGRRYWRGPDSDGQLIFQITAFQEDF-YLHLTP 70
Db 1 ATRPKGAVQPKYEDA--MQEYFKVNGEPVWLHLEKNKG-----LFSDKYSEIHSP 49
QY 71 DAQFLAPAFSTHGLVPLQGLTGGSSDLRCFYSDVNAEPDFAAVALCGLGRCAGFYR 130
Db 50 DGREITTPPVED-----HCYHGRIENDADSTASISACNLGKYFKLQ 93
QY 131 GAQYVITSP--LPNASAPAAQRNSQGAHLQRRVPGSPGDPSTSRGVASGWN----- 181
Db 94 RETYFIEPLKLPDSEAH-----VFKYENVE--KEDEAPKMGVYQNKWSVEPIKK 142
QY 182 -----PAILRALDPYKPRRAGFGESRRSRRAKRFVSPRYVETLVVADESVMKPHG 235
Db 143 ASQALFTAEQQRYPDK-----YIEFFVVDQGTVTTKNN 176
QY 236 ADLEHYLLTLATAARLYRHPSILNPI-NIVVVKVLL--LRDRSGPKVT--GNAALTLR 290
Db 177 GOLD-----KIKARMTYELANIVNEIFLYMHVALVGLWISNGDKITVDPDYTLN 229
QY 291 NFCAQKKLNVSKDHPYWDITAILFTRODLGCGATCTDGLMADYGTMCDPKRSQVIED 350
Db 230 SPAEWRKTDLLTRKKH---DNAQLLTAIDFNG---FTIGVAYIGSMCHPKRSVGIQVD 281
QY 351 ----DGLPAAFTAHGLHVFHNPNDVNVCEVFGKLRANH---MMSPTLQIDRANPWS 404
Db 282 YSPINLVAVIMAHENGLGIHDT-----GSCSGDYPCIMGPT--ISNPSKFFS 332
QY 405 ACSAAITITDLSGHDGDLLOP-----SKPISLPDLPGASVYTLSSOCELAFLGVS 458
Db 333 NCSYIQWDFIMNHPECIINEPLGTDIISPPVCGNELL-----EVGECD----- 378
QY 459 CPYMYQYCKLWCTGKAKGOMVCQTRHFPWADGTCGEGKCLKAGACVERHNLKHRYD-- 516
Db 379 CGTPENCNECCDA-----ATCKLK-----SGSQCGHGD-----CCBQCKFSKGTBCR 422
QY 517 GSAKWADPYGSPRSCGGVQLARRQCTNPTPANGGKYCEGVVRKYRSCNLEPCP----- 571
Db 423 ASMSECDPAEHCTGSS-----ECPADVFKHNGQPCLD---NYGYCYNGNCPIMVHQ 471
QY 572 -----SSASGKSFREQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRD-KC-KLIC 620
Db 472 CYALFGADVYEADSCFDKDNQKGNVYCRKENGKKIICA-----PEDVKCGRLYC 522
QY 621 RANGTG-----YFYVLAPK-----YVDGTLCSPTDSTSVCVGKCI 655
Db 523 KDNSPGQNNPKCMFYSDNDEHKGMVLPCTKCA--DGKVCNSNGHCV 565

RESULT 12

S55270
catrocollastatin precursor - western diamondback rattlesnake
C:Species: Crotalus atrox (western diamondback rattlesnake)
C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C:Accession: S55270
R:Zhou, Q.; Smith, J.B.; Grossman, M.H.
Biochem. J. 307, 411-417, 1995
A:Title: Molecular cloning and expression of catrocollastatin, a snake-venom protein
A:Reference number: S55264; MUID:95251603; PMID:7733877
A:Accession: S55270

A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-609 <2HO>
A:Cross-references: GB:U21003; NID:g710353; PIDN:AAC59672.1; PID:g710354
C:Superfamily: mouse meltrin alpha; disintegrin homology

```
Query Match      9.0%; Score 337; DB 2; Length 609;
Best Local Similarity 22.8%; Pred. No. 3.4e-16;
Matches 150; Conservative 90; Mismatches 259; Indels 160; Gaps 30;

QY 50 DOGLIFQITAFQEDFYHLTPDAQFLAPAFSTEHGLVPLQGLTGSSDLRRCFYSGDNA 109
Db 52 EDAMQVELKNGEPPVHLHKGKGLFSKDYSETHSPDGREITTYPLVEDHCVYHGRIEN 111
QY 110 EPDSFAVSLCGGLRGAFGRGAEYVISP--LPNASAPAAORNSQGAHLQRGVPGPS 167
Db 112 DADSTASISACNKGKHLQGEMLTEPLKLPDSEAHAVKYENVE-----KE 160
QY 168 GPTSRCGVASGNPAILRALDPYKPRRAGFGESESRSSRRSRGAKRFVSIIP-----RY 219
Db 161 DEALKMGVTONW-----ESYEPIK-----KASQLVVTAEHQKYNPFRF 199
QY 220 VETLVVADESMTKFGHDLHY---LLTLATAARLYRHPSTILNPINIVVVKLLLRDRD 276
Db 200 VELFLVVDKAMVTKNNGDLDKIKRMYEIVNTVNEYIRYMY---IHVALVGLIWSNED 255
QY 277 SPKVT--GNAALTNRNCAWOKKLNKVSCKHPEYWDTAILFTRODLGCATTCDTLGMAD 334
Db 256 ---KIIVKPPAGYTLNAGFEWRTDLLTRKH---DNAQLLTAIDL-----DRVITGLAY 303
QY 335 VGMCDPKKSCSVIED---DGLPSAFTTAHELGHVFNMPHDN-VKCEEVFGKLRANHM 390
Db 304 VGSWCHPKRSTGIQDYSEINLVAVIMAHENGHNIGINHDSCYSCGDIACIMRPEISP 363
QY 391 SPITLIQIDRANPSACSAALITDLDGSHGDCLLDQ-----SKPISLPDLPGASYTL 444
Db 364 EPSTF-----FNSGSEFECDWDFIMNHNPECILNEPLGTDIISPPVCGNELL-----EV 411
QY 445 SOCELAFAFGVSGKPCPMYQCYKLTCTGKAKGMVQOTRHFPPWADGTSCGEGKCLKGAC 504
Db 412 GEBGD-----CGTPENCQNECDA-----ATCKLK-----SGSQCGHGD-----C 446
QY 505 VERHNLNKHURVD--GSWAKWDYRPGPCSRTCGGGVOLARQCTNPPTPANGKYGCEGRVKY 562
Db 447 CEQCKFSKSGTECRASMSCECPAEHCTGQSS-----ECPADVFKHNGQPCLD---NY 495
QY 563 RSNCLNPPCP-----SSASGKSFREOCEAFNGYNHSTNRLTLAVANVPKYS 608
Db 496 GYCYNGCNCFIMYHOCYDLFGADVYEAEDSCFERNOKRGNYGYCRKENGKNKIPCA----- 549
QY 609 GVSPRD-KC-KLICRANGTG-----YFYVLAPK-----VVDGTLCSPDSTSVCGQKCI 655
Db 550 ---PEDVKCGRLYCKDNQSPGNPNCKMFYSNEDEHKGMVLPCTKCA--DGKVCNNGHCV 603
```

RESULT 13
S60257

meltrin alpha - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
C:Accession: S60257
R:Yagami-Hironomasa, T.; Sato, T.; Kurisaki, T.; Kamiyo, K.; Nabeshima, Y.I.; Fujisawa-Sel
Nature 377, 652-656, 1995
A:Title: A metalloprotease-disintegrin participating in myoblast fusion.
A:Reference number: S60257; MUID:96026308; PMID:7566181
A:Accession: S60257
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-903 <YAG>
A:Cross-references: EMBL:D50411; NID:g1054586; PIDN:BAA08912.1; PID:g1054587
C:Superfamily: mouse meltrin alpha; disintegrin homology
F:421-503/Domain: disintegrin homology <DIS>
F:349/Active site: Glu #status predicted

Query Match 8.8%; Score 329; DB 2; Length 903;
Best Local Similarity 21.5%; Pred. No. 2e-15;
Matches 159; Conservative 74; Mismatches 245; Indels 262; Gaps 27;

```
QY 2 LLLGLTILTAFAGRTA-----GGFEPEREVVVPIRLDPDINGRRYWRGPEDSGDQG 52
Db 15 LLLAAGALLAPRAARGSLMDQRGAYEVARSL--LSKDPGIPGQSI-----PAKDHPDV 68
QY 53 LIFQITAFQEDFYHLTPDAQFLAPAFSTEHGLVPLQ-----GLTGSSDLRRCFYSGDV 107
Db 69 LTVQLQLESRDLLSLERNEGLIANGFTETH---LQDGTDSLTRNHTD--HCYHGHV 123
QY 108 NAEPSFAVSLCGGLRGAFGRGAEYVISP--LPNAS-----APA-AORNSQGAHLQRG 161
Db 124 QGDAASVYSLTSCDSLRLIMFENKTYSLPEMKNTTDSYKLVPAESMTNIOGL----- 176
QY 162 VFGPSGDPTSRGCVASGNPAILRALDPYKPRRAGFGESESRSSRRSRGAKRFVSIIPRYVE 221
Db 177 -----CGSQHNKSNLTWEDVSP-----GTSQMRARRHK---RETLKMTKYVE 215
QY 222 TLVVADESMTKFGHDLHYLLTLATAARLYRHPSTILNPINIVVVKLLLRDRDSGPKV 281
Db 216 LVIVADNREFQKQKLEKVKQRIEIANHVDKFRPLN-IRIVLGVGVVNDIDK-CSI 273
QY 282 TGNAAITLRNCAWOKKLNKVSCKHPEYWDTAILFTRODLGCATTCDTLGMADYGTMDP 341
Db 274 SODPTRLHFEFLWRKIKLPRKSH---DNAQLISGYVFOGT-----TIGNAPIMSMCTA 325
QY 342 KRSCSVI---EDDGLPSAFTTAHELGHVFNMPHDN-VKCEEVFGKLRANHMSPTLIQI 397
Db 326 EQSGGVVMDHSDSPLAGAVTLAHELGHVFNHHTLERGSCRMARAEKGGCIMPSS----- 381
QY 398 DRANP---WSACSAALITDLDGSHGDCLLDQ-----SKPISLPDLPGAS----- 427
Db 382 -TGFPFVVFSSCKRDLKLEASLEKGMGMCNLPNPEVQAFQAGKGGNGYVEEGECDCGE 440
QY 428 -----SKPISLPDLPGAS-----SKPISLPDLPGAS----- 441
Db 441 PEECTNRCCNATTCTLPDAVCAHQCCEDCQLKPPGTACRGSSNSCDLPEFCTGTAPHC 500
QY 442 -----YTLSSQCELAFAFGVSGKPCPMYQY----- 464
Db 501 PANVYLDHGHPCQGVGYNGYICQTHEQQCVTLWGFKAPEAGICFERVNSAGDPYGN 560
QY 465 -----CTKLWCTGKAK-----GOMVQOTRHPFW 487
Db 561 GKDSKSAFAKCELDAKCGKIQCGQGSAPVIGTNAVSIETNIPQEGGRILCRGTHVYL 620
QY 488 AD-----GTSCGEGKCLKGACVERHNLNKHURVDGWSWAKWDYRPGCSRTC-GGV 536
Db 621 GDDMPDCLVLAGTKCAEGKICLNRRQNISVFGVHK-----CAMOCHGRGV 667
QY 537 QLARRQCTNPPTPANGKYCE 556
Db 668 CNNRKNC-----HCE 677
```

RESULT 14

S48160
metalloproteinase (EC 3.4.24.-) H-I precursor - carpet viper
N:Contains: disintegrin
C:Species: Echis pyramidum leakeyi
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C:Accession: S48160
R:Palme, M.J.I.; Moura-Da-Silva, A.M.; Theakston, R.D.G.; Crampton, J.M.
Eur. J. Biochem. 224, 483-488, 1994
A:Title: Cloning of metalloprotease genes in the carpet viper (Echis pyramidum leakeyi)
A:Reference number: S48160; MUID:95010025; PMID:7925363
A:Accession: S48160
A:Molecule type: mRNA
A:Residues: 1-617 <PAI>
A:Cross-references: GB:X78970; NID:g763092; PIDN:CAA55565.1; PID:g763093

C:Superfamily: mouse meltrin alpha; disintegrin homology
C:Keywords: hydrolase; metalloproteinase; venom
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-192/Domain: propeptide #status predicted <PRO>
F:193-617/Product: metalloproteinase H-II #status predicted <MAT>
F:403-485/Domain: disintegrin homology <DIS>
F:337/Active site: Glu #status predicted

```

Query Match      8.6%; Score 323.5; DB 2; Length 617;
Best Local Similarity 23.2%; Pred. No. 3.1e-15;
Matches 150; Conservative 74; Mismatches 267; Indels 155; Gaps 27;

QY 45 PEDSGQGLIFQITAFQEDFYHLTPDAQFLAPAFSTHGLVPLQGLGGSSDLRRCRYS 104
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
DB 50 PEQYEDTMRVEKNGEPEVWLEKNGLFSEDIYSETHSPDGREITNPPVEDHCYH 109
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
QY 105 GOVNAEPDSFAVSLCGGLRGAFGYGAEYVISPL--PNASAPAAQRNSQGAHLRQGV 162
   | : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 110 GRIQDADSTASACNGLKGFLMRLGETYLIEPLKIPDSEAHAYKYENVE----- 161
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 163 PGSPGSDPTSRGVA--SGWNPAILRALDPYKPRRAGFESRRSRRAKREVSIPRYVE 221
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 162 ---KEDEAPKMGVGTNNES-----DELKASQIVATSEQR-----SYKKIE 203
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 222 TLVVADESMVKFGADLEHYLLTLATAARLYRHPISILNPINIV-----VKVLLR 273
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 204 FVVVADYIMRYKNND-----STAVRRRIEIVNLMVMVIVFNHVALTHIEIWSTR 256
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 274 DRSGPKYVGNALTLRNFCAQKLNKYSKHPEDWDFAILFTRODLGGATTCDTLGN 333
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 257 DQIT---VOSAADVTLDFGDMRAKNLLTRKKH---DNAQLFTGINLNG----- 305
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 334 DVGTMDCKPRKSCSVID---DGLPSAFTTAHELGHVFNPHDNVNVK--CEVEFGKLRANHM 389
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 306 RMKGMSPNSVGVQIDYCKNVLVVAITMAHELGNLGHNDGNCNCPDTSICINSAVAG 365
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 390 MSTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPSLPELPGASYT--LSQOC 448
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 366 PEPVF-----SFSNCRNDYRSFRNSDQSKIDNKLKLTQIVSPVCGNVEVEGEC 418
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 449 ELAFVGSKPCPYMQYCTKLWCTGKAKGMQVQTRHFPWADGTSCGEGKCLKGACVERH 508
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 419 D-----CGSRVTCRNPCCNA-----TTCKL-----TPGQCADGEC--NCRFRP 457
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 509 NLNKR-----VDGSAKWDPYGPCSRTC--GGVQLARRQTN--545
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 458 ARTECRKIDDCDPEYCTGQSECPDVFQNGQPYQSNNGCYNGNCPILKNQCIHLW 517
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 546 -PTPANG-----GKYCEGVVRYKRSNLEPCPSSASGKSF 579
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 518 KPAPAGVNVAPDVCFDNQNGTDVYVNCGIKNGTYIKCARQDIK--CGRLSCEEPSTGNTI 576
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 580 REQCEAFNGYHNHSTNRLTLAVAWPKYSGVSPRDKCK--LTCRAN 623
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 577 ---NCAFFS---SRN-----PDYGMVDIGTKCADGKVCNSN 607
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :

```

RESULT 15
S48169
N:Contains: disintegrin
C:Species: Echis pyramidium leakeyi
C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 09-Jun-2000
C:Accession: S48169
R:Paiva, M.J.I.; Moura-Da-Silva, A.M.; Theakston, R.D.G.; Crampton, J.M.
Eur. J. Biochem. 224, 483-488, 1994
A:Title: Cloning of metalloproteinase genes in the carpet viper (Echis pyramidium leakeyi).
A:Reference number: S48169; MUID:95010025; PMID:7925363
A:Molecule type: mRNA
A:Residues: 1-549 <PAI>
A:Cross-references: EMBL:X78971; NID:g763094; PIDN:CAA55566.1; PID:g763095
C:Superfamily: mouse meltrin alpha; disintegrin homology

C:Keywords: hydrolase; metalloproteinase; venom; zinc
F:1-138/Domain: propeptide (fragment) #status predicted <PRO>
F:139-549/Product: metalloproteinase H-II #status predicted <MAT>
F:350-432/Domain: disintegrin homology <DIS>
F:284/Active site: Glu #status predicted

```

Query Match      8.4%; Score 313; DB 2; Length 549;
Best Local Similarity 23.7%; Pred. No. 1.5e-14;
Matches 156; Conservative 77; Mismatches 234; Indels 192; Gaps 35;

QY 62 EDFVLHLTPDAQFLAPAFSTHGLVPLQGLTGGSSDLRRCFYSGDVNAEPDSFAAVSLCG 121
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 12 EPVVLHLOKNDLESEDIYSETHYTPDGREITNPPVEDHCYIYHGRVONDHAHSASISACN 71
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 122 GLRGAFGYRGAEEYVISPL--PNASAPAA-----QRNSQGAHLRQRRVPGSPGDPTSR 173
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 72 GLKGHFKLOGETYFIEPLKIPDSEAHAYKYENIEKEDQAPK-----114
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 174 CGVA--SGWNPAILRALDPYKPRRAGFESRRSRRAKRFVSIPIRYVETLVVADSWVK 232
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 115 CGVTHTNW-----ESDEPIK-----EASRLVASSEQQSYNDNFYVKLVIVVDHRMVT 162
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 233 FHGAD--LEHYLLTLATAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALTLR 290
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 163 KYNDSTIRTRYEMVNTVNEIYH---LN--IRVALGLEFWSNGDL--INVTSAEHTLN 217
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 291 NFCAWQKLNKVS-----RHPEYWDTAILFTRODLGGATTCDTLGMADYGTMDCKPRSC 345
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 218 LFGVW-----RASDLLSRRRH---DNAHLTLAIDLNG---PTIGLRDYSSMCQATRSV 264
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 346 SVIEDDLGPS-----AFTTAHELGHVFNPHDNVNVKCEEVFGKLRANHMSPFTLIQIDRAN 401
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 265 GWVQDHS--PTVRVAVTMAHEMGNLGNSHDG--NHC-----NCGANSCIMAAVLRNPAPE 317
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 402 PWSACSAIITDFLDGSHGDCLLDQPSK--PISLPEDLPGASYTLSSQOCELAFVGVSKPCP 460
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 318 YFDCSRRYQNFNTNIPDCTIIRPSKTDIVSPQVCGNGLLDEGECD-----366
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 461 YMQYCTKLWCTGKAKGMQVQTRHFPWADGTSCGEGKCLKGACVERHNLNKHVDSGWA 520
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 367 -----CGSPAN-----CQ---YPCDDAASC--KL-----HSWV 389
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 521 KWDYGPGRSRTCGGVQLARRQCTNPTPANGGKYCEGVVRYKRSNLEPCPSSASGKSF 580
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 390 ECE--FGHCDDQC-----RFPKA--GTECRGIR---SECDL---PEYCTGQS--426
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 581 EEQCEAFNGYHNHSTNRLTLAVAW---VPKY-----SGVSPRDKC-----616
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 427 AECLMEFSTRWKTWPKLSLLLOWTCPIWQYCYAHFGONAVVGODACFEINKEGKGFYC 486
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
QY 617 -----KLICRA--NGTGYFYVLA PKVVDGTLCLSPDSTSVGVQKCI 655
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 487 RKENDVPIPCAQEDVKCGRLFCETEPNNMCRYPYDEGMVDPGTRC--EDKVKCINGKCI 543
   || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :

```

Search completed: April 29, 2003, 17:13:30
Job time : 20.8904 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 8.4611 Seconds
(without alignments)
3367.676 Million.cell updates/sec

Title: US-10-009-332-1-copy_1_687

Perfect score: 3744

Sequence: 1 MLLGLITLAFAGTAGGFE.....RFDKGVCGDNKCKKVTG 687

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1960.5	52.4	967	1	ATSL_HUMAN	Q9uh18	homo sapien
2	1947.5	52.0	968	1	ATSL_MOUSE	P37857	mus musculus
3	1934	51.7	967	1	ATSL_RAT	Q9wuq1	rattus norv
4	1639.5	45.4	890	1	ATSL_HUMAN	Q9up79	homo sapien
5	1689.5	45.1	837	1	ATSL_HUMAN	Q75173	homo sapien
6	1679	44.8	905	1	ATSL_MOUSE	P57110	mus musculus
7	1613	43.1	930	1	ATSL_MOUSE	Q9r001	mus musculus
8	1587.5	42.4	930	1	ATSL_HUMAN	Q9una0	homo sapien
9	1543.5	41.2	1629	1	ATSL_HUMAN	Q9p2n4	homo sapien
10	1518.5	40.6	630	1	ATSL_RAT	Q9esp7	rattus norv
11	1087	29.0	997	1	ATSL_HUMAN	Q9uk54	homo sapien
12	1072.5	28.6	1077	1	ATSL_HUMAN	Q9h324	homo sapien
13	1023	27.3	1593	1	ATSL_HUMAN	P58397	homo sapien
14	994	26.5	1211	1	ATSL_HUMAN	Q95450	h adamts-2
15	980.5	26.2	1205	1	ATSL_BOVIN	P79331	b adamts-2
16	968.5	25.9	1205	1	ATSL_HUMAN	O15072	homo sapien
17	918.5	24.5	860	1	ATSL_HUMAN	Q9ukp5	homo sapien
18	601.5	16.1	245	1	ATSL_BOVIN	Q9tt93	bos taurus
19	599	16.0	207	1	ATSL_BOVIN	Q9tt92	bos taurus
20	366.5	9.8	824	1	AD08_HUMAN	P78325	homo sapien
21	360.5	9.6	776	1	AD28_MACFA	Q9xsl6	macaca fasc
22	346.5	9.3	813	1	AD33_HUMAN	Q9bz11	homo sapien
23	344.5	9.2	571	1	DIJ1_BOVJA	P30431	bothrops ja
24	342.5	9.1	774	1	AD28_MOUSE	Q9iln6	mus musculus
25	336.5	9.0	775	1	AD28_HUMAN	Q9ukq2	homo sapien
26	331.5	8.9	956	1	AD19_HUMAN	Q9h013	homo sapien
27	329	8.8	903	1	AD12_MOUSE	Q61824	mus musculus
28	325.5	8.7	920	1	AD19_MOUSE	Q35674	mus musculus
29	323	8.6	857	1	AD22_MOUSE	Q9rlv6	mus musculus
30	315.5	8.4	906	1	AD22_HUMAN	Q9p0k1	homo sapien
31	314	8.4	909	1	AD12_HUMAN	QK3184	homo sapien
32	313	8.4	826	1	AD08_MOUSE	Q05910	mus musculus
33	307.5	8.2	789	1	AD07_RAT	Q63180	rattus norv

34 297 7.9 480 1 DISA_TRIGA p15503 trimeresuru
35 296 7.9 413 1 ACIA_ACKAC Q9pw35 agkistrodon
36 288 7.7 788 1 AD07_MOUSE Q35227 mus musculus
37 284 7.6 414 1 HRD7_CROAT P15167 crotalus at
38 279.5 7.5 769 1 AD11_HUMAN Q75078 homo sapien
39 277.5 7.4 814 1 AD15_HUMAN Q13444 homo sapien
40 273.5 7.3 478 1 HRTE_CROAT P34182 crotalus at
41 270.5 7.2 773 1 AD11_MOUSE Q9rlv4 mus musculus
42 268 7.2 797 1 AD33_MOUSE Q923w9 mus musculus
43 260.5 7.0 754 1 AD07_HUMAN Q9h2u9 homo sapien
44 260.5 7.0 935 1 AD22_XENLA Q42596 xenopus lae
45 257 6.9 776 1 AD07_MACFA Q28475 macaca fasc

ALIGNMENTS

RESULT 1
ATSL_HUMAN
ID ATSL_HUMAN STANDARD; PRT; 967 AA.
AC Q9UH18; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
GN ADAMTS1 OR METH1 OR KIAA1346.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
RT "Cloning, characterization and mapping on human chromosome 21 of the
orthologue of murine Adamts-1.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC TISSUE=Heart;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
Lombardo M., Trueta-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
family of proteins with angio-inhibitory activity.";
RL J. Biol. Chem. 274:23349-23357(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=20247184; PubMed=10785405;
RA Glienke J., Schmitt A.O., Pillarsky C., Hinzmann B., Weiss B.,
Rosenthal A., Thierauch K.H.;
RT "Differential gene expression by endothelial cells in distinct
angiogenic states.";
RL Eur. J. Biochem. 267:2820-2830(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
The complete sequences of 150 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,


```
QY 207 -----SGRAKREYSPRYVETLVVADSMVKFHGADLEHYLLTLLATAARLYRHPHSILNPI 262
DQ 244 SGPFSIRKKREYSSPRYVETLMVADQSMADFHGSLGKHYLLTLESVAARFYKHPISIRNSI 303
QY 263 NIVVVKVLLLRDRDQSPKVTGNAALTLRNFCAWQKLNKVSQKHPYWDTAILFTRQDLC 322
DQ 304 SLVVVKILVIEYEQKGPVTSNAALTLRNFCAWQKLNKVSQKHPYWDTAILFTRQDLC 363
QY 323 GATTCDTLGMADVGTWCDPKRSCSVTEDDGLPSAFTTAHELGHVFNMPHDNVKVCCEVFG 382
DQ 364 GSHTCDTLGMADVGTWCDPKRSCSVTEDDGLPSAFTTAHELGHVFNMPHDNVKVCCEVFG 423
QY 383 KLRANHMSPTLIQIDRANPWSACSAAITDFLDHSGHGLDQSPKISLPEDLPAGASY 442
DQ 424 VTGDSHLMASMLSSLDHSPWSPVSMYVTSFLDNGHGBCLMDKPNKIPKPSDLPGLTLY 483
QY 443 TLSQCELAFLGVGSKCP-PMOYCTIKLWCTGKAGKGMWCOTRHFPAWDTSGEGKLCIK 501
DQ 484 DANRQCOFTFGEESKCPDAASTCTTLWCTGTSGGLLVQCTKHFPWADTSGEGKWCVS 543
QY 502 GACVERHNLNKH---RVDGSAKWDYPGCSRTCGGGVQLARRQCTNPTTPANGGKYCEGV 558
DQ 544 GKCVNKTDM-KHPATPVHSGWGPWGPWDCSRTCGGGVQYTMRECDNPVPKNGKYCEGK 602
QY 559 RVYRSCNLEPCPSSASGSKSPREEQCEAFNGYNHSTNRLTLAVAVPKYSGVSPRDKCL 618
DQ 603 RVYRSCNLEPCPSSASGSKSPREEQCEAFNGYNHSTNRLTLAVAVPKYSGVSPRDKCL 661
QY 619 ICRANGTVFYVLAPKVPDGTCLSPDSTSVCGVGKIKAGCDGNLGGSKKPKDKCGVCGGD 678
DQ 662 TCAKAGIGFVFLQPKVVDGTCPDSTSVCGVGKIKAGCDGNLGGSKKPKDKCGVCGGD 721
QY 679 NKSCCKKVTG 687
DQ 722 GSTCKKMSG 730

RESULT 3
ATSI_RAT
ID ATSI_RAT STANDARD; PRT; 967 AA.
AC 09WU01; 09ERIL;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RT "Induction of a disintegrin and metalloproteinase with the
RT thrombospondin type I motif (ADAMTS).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 18-967 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luethi M., Hoesli M., Reichen J.;
RT "Cloning of the rat ADAMTS-1 gene and its down regulation in
RT endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
```

```
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CC CIRRHOTIC LIVER.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B..
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF149118; AAD34012.1; -
CC EMBL; AF304446; AAG29823.1; -
CC MEROPS; M12.222; -
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_MTPeptidse.
CC Pfam; PF00090; tsp_1; 6.
CC Pfam; PF01421; Reprolysin; 2.
CC Pfam; PF01562; Pep_M12B_propep; 2.
CC SMART; SM00209; TSP1; 3.
CC PROSITE; PS00215; ADAM_MEPRO; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS00092; TSP1; 2.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix; Heparin-binding.
CC SIGNAL 1 54 POTENTIAL.
CC PROPEP 55 252 BY SIMILARITY.
CC CHAIN 253 967 ADAMTS-1.
CC SITE 205 205 CYSTEINE SWITCH (POTENTIAL).
CC METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
CC ACT_SITE 402 402 BY SIMILARITY.
CC METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
CC DOMAIN 476 558 DISINTEGRIN-LIKE.
CC DOMAIN 559 615 TSP TYPE-1 1.
CC DOMAIN 616 724 SPACER.
CC DOMAIN 725 857 CYS-RICH.
CC DOMAIN 858 907 TSP TYPE-1 2.
CC DOMAIN 908 967 TSP TYPE-1 3.
CC DOMAIN 194 198 POLY-ARG.
CC CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 21 21 I -> V (IN REF. 2).
CC CONFLICT 26 31 KFRSQ -> RSRGSL (IN REF. 2).
CC CONFLICT 49 49 V -> A (IN REF. 2).
CC CONFLICT 72 72 R -> P (IN REF. 2).
CC CONFLICT 79 79 L -> TR (IN REF. 2).
CC CONFLICT 249 249 R -> G (IN REF. 2).
CC CONFLICT 262 265 TMLV -> NLK (IN REF. 2).
CC CONFLICT 607 607 S -> F (IN REF. 2).
CC CONFLICT 936 936 L -> V (IN REF. 2).
CC CONFLICT 962 962 I -> T (IN REF. 2).
CC SEQUENCE 967 AA; 105705 MW; F93C864F6DCDB4CF CRC64;
```

Query Match 51.7%; Score 1934; DB 1; Length 967;
Best Local Similarity 52.7%; Pred. NO. 2.8e-136;
Matches 378; Conservative 98; Mismatches 187; Indels 54; Gaps 15;

QY 1 MLLGLTILAFAGRTAGG--FPEREVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQIT 58
DB 37 LLLLSITMLLCVGAHGRTPEDEBELV-----PSLRARGH-----DSTVLLRLD 83
QY 59 AFQEDFVHLTDAFAPAFSTHGLVPLGLTGS-----SDLRRFCYSDVNAE 110
DB 84 AFQQLHLKLQDSEGLAPGFTLV-----GRSPGSAQHLDPDGLAHCFYSVWGD 138
QY 111 PDSFAVSCGLRGAFGYRGAEYVISPPLNAS-----APAAQNSQGA-----HLI--QRR 160
DB 139 PSSAALSCEGVAFYGLQEEFTIQAPAVATERLVPAPKEESIAAPRPHILRRRR 198
QY 161 GVPGPSGD-----PTSCGVASGWNPAILRALDPYPRAGFESRRSRRAKRFV 214
DB 199 GSGGAKCGVMEETLPTNSGRESQNTPDWPLRNP--TPQAG-----KPTGPGSIRKRFV 254
QY 215 SIPRVETLVWADSNMFKHGADLEHLLTLATARLYRHSILNPNINVVVKVLLRD 274
DB 255 SPFRVETMLVADQSMADFHGSLGLYLLTFVSAARYKHSIRNSISLVVKVLYIE 314
QY 275 RDSGPKVTGNAALTRNFCAMOKLNKVDKHPYWDTAILFTRODLGCATTCDTLGMD 334
DB 315 EQKGPEVTSNAALTRNFCAMOKLNKVDKHPYWDTAILFTRODLGCATTCDTLGMD 374
QY 335 VETMCDPRKSCVIEDDGLPSAFTAHGLGHVFNPHNVKVCVEVFKGLRANHMSP 394
DB 375 VGTVCDPSCVIEDDGLQAAFTAHGLGHVFNPHNDKACFNSGVSGDHLMA 434
QY 395 IOIDRANPWSACSAITLDFDSHGDCILLOPKSPILPDLPCASYTLSSQCELAGV 454
DB 435 SLDSHSQSPSCSAVYVTFDNGHGECIMDKPQNPILPDLPTLYDANRCQFTFE 494
QY 455 GSKPCP-YMOYCTKLWCTGKAKGMVQCOTRHPFWADGTSCGEGKLCCLKGACVERHN 513
DB 495 ESTHCPDAASTCTLWCTGSLGLVCTCKHEPWADGTSCGEGKWCVCVKCNKTD 553
QY 514 ---RVDGSAKWDPYGPSCRTGGGVQVLAARCTNPTFANGKCYEGVVRVYRSCNLEPC 570
DB 554 FATPVHSGWPGMPCDSCRTGGGVQVYTMRECDNPVKNKGKCYEGKRVYRSCNIEDC 613
QY 571 PSSAGKSPREOCFAFNGYHNSTNRLTLAVAWPKYVSVSPDKCKLICRANGTGYEV 630
DB 614 PNN-NKGTFRQCAHNEFSASFGNEPTVETPKYAGVSPDKCKLICRANGTGYEV 672
QY 631 LAPKVVDTLCSPDSTSVQVKCIKAGCDGNLGSKKRFRDGVCGGDNKSKCKVTG 687
DB 673 LQPKVVDGTPCSPDSTSVQVQCVKAGCDRIIDSKKFRDGVCGGNGSTCKKISG 729

RESULT 4
ID AT58_HUMAN STANDARD; PRT; 890 AA.
AC Q9UP79; Q9NZSO;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2) (METH-8).
DE (METH-8).
GN ADAMTS8 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,

RA Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity.",
RL J. Biol. Chem. 274:23349-23357(1999).
RN [2]
RP SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on mouse chromosome 9 and human chromosome 11.",
RL Genomics 62:312-315(1999).
CC -1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND KIDNEY.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF060153; AAD48081.1; -
CC EMBL: AF175283; AAP25806.1; -
CC HSP: P34179; IIAQ.
CC MEROPS: M12.226; -
CC Genew: HGNC:224; ADAMTS8.
CC MIM: 605175; -
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Reprylisin.
CC InterPro: IPR000884; TSP1.
CC InterPro: IPR000130; Zn_MTPpeptdse.
CC Pfam: PF00090; tsp_1; 2.
CC Pfam: PF01421; Reprolysin; 1.
CC Pfam: PF01562; Pep_M12B_propep; 1.
CC SMART: SM00209; TSP1; 2.
CC PROSITE: PS00215; ADAM_MEPRO; 1.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC PROSITE: PS50092; TSP1; 1.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL; 1 27 POTENTIAL
FT PROPEP 28 214 BY SIMILARITY.
FT CHAIN 215 890 ADAMTS-8.
FT METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 365 365 BY SIMILARITY.
FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 439 526 DISINTEGRIN-LIKE.
FT DOMAIN 527 583 TSP TYPE-1 1.
FT DOMAIN 584 690 CYS-RICH.
FT DOMAIN 691 832 SPACER.
FT DOMAIN 833 890 TSP TYPE-1 2.
FT DOMAIN 202 205 POLY-PRO.
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 600 600 N-LINKED (GLCNAC. . .) (POTENTIAL).


```
DR MM: 603876;
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSPl.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF00090; tsp_1; 1.
DR Pfam: PF01421; Reprolysin; 1.
DR SMART: SM00209; TSPl; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS0215; ADAM_MEPRO; 1.
DR PROSITE: PS50092; TSPl; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Extracellular matrix.
FT SIGNAL 1 51
FT PROPEP 52 212 POTENTIAL.
FT CHAIN 213 837
FT SITE 194 194 ADAMTS-4.
FT METAL 361 361 CYSTEINE SWITCH (POTENTIAL).
FT ACT_SITE 362 362 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 365 365 BY SIMILARITY.
FT METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 437 519 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 520 576 DISINTEGRIN-LIKE.
FT DOMAIN 577 685 TSP TYPE-1.
FT DOMAIN 686 837 CYS-RICH.
FT DOMAIN 837 837 SPACER.
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 77 77 A -> T (IN REF. 1).
FT CONFLICT 626 626 R -> Q (IN REF. 3).
FT CONFLICT 682 682 G -> R (IN REF. 3).
FT SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;

Query Match
Best Local Similarity 45.1%; Score 1689.5; DB 1; Length 837;
Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;

QY 1 MLLGILLIAPAGRTAGFEPEVIVVPIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
DB 37 LLLLLASLLSARLASPLPEEEIVFPEKLNGSVL-----PGSGAPARLLCRLQAF 88
QY 61 QEDFVHLTPDPAFLAPAFSTHGLVPLQGLTGSSDLRRCFYSDVNAEPDPSFAAAYSLC 120
DB 89 GETLLELEQDSGVQVEGLTVQLQAPF--LLGAEF--GTYLTGTNDGPESVASLHWD 145
QY 121 GG-LRGAFGYGAEVYISPLPNASAPAAQRNSOGAHLQRRVGPSPGSDPSTRCGVASG 179
DB 146 GCALLGVLYRGAELHLPLEGGTPNSA--GGPGAHLRRK-----SPASGGPNCNV--- 196
QY 180 WNPATLRALDPKPRACFGESRRSRGRKRFRVSPRYVETLVVADESVMVKFHGADLE 239
DB 197 -----KAPLGSFSPRPR--RAKRFASLSRFVETLVVADDKMAAFHAGLKL 239
QY 240 HYLTLTATAARLYRHPSILNPINIVVVKVLLRRDRSGPKVTCGNAALTRNFCWAKKL 299
DB 240 RYLLTVMAAAKAKHPISIRNPVSLVTRVILSGEGPGVGFSAQTLSFCWARGQL 299
QY 300 NKVSDKHPYNDTAILFTRODLGGATCDTLGMADVGTMCDPKRSCSVIEDGLPSAFTT 359
DB 300 NTPEDSDPDHFDTAILFTRODLGGVSTCDTLGMADVGTVCDPARSCAIVEDDGLQSAFTA 359
QY 360 AHELGHVFNMPHDNVKVEEYFGKL--RANHMSFTLQIDRANPWSACSAAITDFDLDSG 418
DB 360 AHELGHVFNMLHDSKPCISLNGPLSTSRHYMAPVMAHVDPPEEPWSPCSARFITDFDLNG 419
QY 419 HGDCILLDQSPKIPISLPEDLPASVTLSSQCELAFGVSKPCPYM--QYCTKLWCTGKAKQ 477
DB 420 YGHCLLDKPEAPLHPVTFPGKDYADQCOLTEGPDNRHCPQPPCCALWCSSGHLNGH 479
QY 478 MYCOTRHPFWADGTSCGEGKLCACVERHNLNKRVD--GSWAKWDYPGCPGSRCTCGG 535
DB 480 AMCQTKHSPWADGTPCGPAQACMGRCILHMDQLQDENIPQAGGVPWGPWGDCSRCTCGG 539
```

```
QY 536 VOLARROCTNPTPANGKYCEGVVRVYKYSRNLNCPSPSSASGSKSFRBQCEAFNGYNHSTN 595
DB 540 VQFSRCDCTRPVPRNGKYCEGRTRFRSCNTEDCP--TGSALTFRBQCAA---YNHRTD 595
QY 596 RLTL---AVAWVPKYSGVSPDKCKLICRANGTGYFYVLAPKVVVDGTLCSPDSTSVQVQ 652
DB 596 LFKSFPGMDWVPRYTGVAPODQCKLTQARALGYVYVLEPRVVDGTPCSPDSSSVQVQ 655
QY 653 KCIRAGCDGNLGSKKRFDKCGVCGGDNKSKKVTG 687
DB 656 RCIHAGCDRIIGSKKFKDKMVGCGDGGSGCKSQSG 690

RESULT 6
ATS8_MOUSE
ID ATS8_MOUSE STANDARD; PRT; 905 AA.
AC P57110;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2).
GN ADAMTS8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID-10090;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE-20079168; PubMed-10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
RT mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC -1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART
CC AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF175282; IAAF25805.1;
DR HSP: P34179; IIAF.
DR MEROPS: M12.226;
DR MGD: MGI:1353468; Adamts8.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSPl.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF00090; tsp_1; 2.
DR Pfam: PF01421; Reprolysin; 1.
DR SMART: SM00209; TSPl; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS0215; ADAM_MEPRO; 1.
DR PROSITE: PS50092; TSPl; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
```

FT	SIGNAL	1	28	POTENTIAL.
FT	PROPEP	29	228	BY SIMILARITY.
FT	CHAIN	229	905	ADAMTS-8.
FT	METAL	378	378	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	379	379	BY SIMILARITY.
FT	METAL	382	382	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	388	388	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DOMAIN	453	541	DISINTEGRIN-LIKE.
FT	DOMAIN	542	598	TSP TYPE-1 1.
FT	DOMAIN	599	705	CYS-RICH.
FT	DOMAIN	706	847	SPACER.
FT	DOMAIN	848	905	TSP TYPE-1 2.
FT	CARBOHYD	415	415	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	480	480	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	506	506	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	905 AA:	99879 MR;	124D4132B33A0CAE CRC64;
Query Match 44.8%; Score 1679; DB 1; Length 905;				
Best Local Similarity 46.7%; Pred. No. 2.4e-117;				
Matches 344; Conservative 101; Mismatches 202; Indels 90; Gaps 18;				
Qy	2	LLLGILRL-----AFAGRTAGGPEPEREVVPIRLDPDINGRYVNRGPDSDGQGL	53	
Db	13	LLLLLLQPPPLVCGAPAGGTGA--QASELVVPTRL-----PGSAFE--L	55	
Qy	54	IFQITAFQDFYHLTDAQFLAPAFSTEHLGVPLQGLTGGSDLRRCFYSGDVNAEPDS	113	
Db	56	AFHLSAFQGVFLRLAPDASFLAPEFKIERLG--GSSAAAGGEPGLRCGFSGVNGERES	114	
Qy	114	FAAYSCGGLRGARGYGAEVVISPLNAPAAQNSQGA-----HLLQRGVPGGP	166	
Db	115	LAAMSCVAGWSGSLLAGEEFTIQP-----QAGDSLDPHRLQRWG-PGOR	160	
Qy	167	SCDPTSCRGVAGSNPAIRLALDPYKPRRAGFGEERS-----RRSG-----	208	
Db	161	REDP-----GLAAAEVFLPQGLWEVEMGNGQGOERDNEEDRKQDKGLLKETE	216	
Qy	209	-----RAKRFYSIPRYVETLVVADPMVKFAGDLHLTLTATAARLYRHPISLN	260	
Db	217	PPPGSKTRSRFFSEARFVETLLVADASMAAFVGTDLQNHILTVMSAARIKHPISRN	276	
Qy	261	PINIVVVKLLLRDRSDGPKVYTGNAALTLRNFCAMQKLLKNSDKHPYWDTAILFTQD	320	
Db	277	SVNLVVKVLIIVERKMGPEVSDNGGLTLRNFCQWRFNKPSDRHPEDHYDTAILFTRQ	336	
Qy	321	LCG-ATTCDTLGMADVGTCDPKBSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKKVCEE	379	
Db	337	FCGKGEQCDTLGMADVGTICDPPDKSCSVIKDEGLQAAYTLAELGHVLSMPHDDSKPCVR	396	
Qy	380	VFGKLRANHMSPTLIQIDRANPWSACSAITDFLDSGHGDCILLDQPSKPSISLPDLP	439	
Db	397	LFGPMGKYHMAFFFIHVNTKLPWSPCSAVYLTLLDDGGHGDCLLDAPTSVLPLTGLP	456	
Qy	440	AS--YTLSSQCELAGVSGKCP--YMOYCTKLWCTKGAKGQWVCOTRH--FPWADGTS	492	
Db	457	HSTLYELDQCKQIFGPDFRHCNPTSVEDICVQLCARHRSDEPICHTKNGSLWADGTP	516	
Qy	493	CGEGLCKLKGACVBRHNLN--KRVDSWAKWDYPCSRCTCGGVQLARRQCTNPTNPAN	550	
Db	517	CGPHLCLDGSCLVKEDVENPKAVVDGNGWPRFWGQCSRTCGGIGQFSNRECDNPNPQN	576	
Qy	551	GKYGCEGVRYKRYSCNLEPCPSSASGSKSFREEQCEAFNGYHNHSTNRLTLAVAWPKYSGV	610	
Db	577	GGRFCLGERVYQSCNTEECF--PNGSFRQCEQCEKYNAYNH--TDLGNFLQWVPKYSGV	633	
Qy	611	SPRDKCKLICRANGTYFYVLAPKVVDGTLTCLSPDSTSCVCGKCIKAGCGNLGSKKREF	670	
Db	634	SPRDKCKLCFRANGSEKFEAKVIDGTLCGPDTLSTICVRGQCQVACGCDHVNSPKRLD	693	
Qy	671	KGVCGGDNKSKKVTG 687		
Db	694	KGVCGGKGTACKRISG 710		

RESULT 7
ID AT55_MOUSE STANDARD; PRT; 930 AA.
AC 09R001;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2) (ADMP-2) (Implantin).
DE ADAMTS5.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.; PubMed=10464288;
RX MEDLINE=99395124; Hirschfeld S., Seidman M.F., Apte S.S.;
RA Hirschfeld S., Apte S.S., and ADAM-TS7, novel members of a new family of
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT zinc metalloproteinases.";
RT J. Biol. Chem. 274:25555-25563(1999).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393 site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR UNDETECTABLE LEVEL THEREAFTER.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF140673; AAD56356.1; -
DR MEROPS; M12.225; -
DR MGD; MGI:1346321; Adamts5.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR00130; zn_M12B_propep.
DR Pfam; PF00090; tsp.1; 2.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00442; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN; 1; FALSE NEG.
KW hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 261
FT CHAIN 262 930 ADAMTS-5.

```
FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 411 411 BY SIMILARITY.
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 566 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE-1 1.
FT DOMAIN 624 731 CYS-RICH.
FT DOMAIN 732 874 SPACER.
FT DOMAIN 875 930 TSP TYPE-1 2.
FT DOMAIN 41 46 POLY-ALA.
FT DOMAIN 257 261 POLY-ARG.
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 930 AA; 101780 MW; 84DE84B26170D4DC CRC64;

Query Match 43.1%; Score 1613; DB 1; Length 930;
Best Local Similarity 47.4%; Pred. No. 2.1e-112;
Matches 306; Conservative 81; Mismatches 182; Indels 76; Gaps 7;

QY 90 GLTGSSDLRCFYSGDVNAEPDFAVSLCGGLRGAFYRGAEYVISPPLNASAPAAQR 149
DB 121 GLSASSHRGHCFYRGVTDGSPRSLAVFDLCGLDGFYAVKHARYTLKPLRSGWAEYR 180
QY 150 -----NSOAHLLORRG-----VPGPGSGDPTSCRGVAGSNWPAILRALDPY 191
DB 181 IYDGSRRILHYNREGFSFALPRASCETPASPFG----- 217
QY 192 KPRRAGFSGSRRRSRG-----RAKRFVSIPIRYVELTVAD 227
DB 218 -POESPSVHSRRSRASALAPOLLDSAFSPSGNAGPQTWRRRRRSISRARQVELLVAD 276
QY 228 ESMVKFHGADLEHYLLTLTATAARLYRPSILNINIVVKKVLLRRDSDGPKVTGNAAL 287
DB 277 SSMRWYRGQHQHLLTLASTANLYSHASIEHRLAVKVVVLTDKDTSLEYSKNAAT 336
QY 288 TLRNFCAMQKLNKVSOKHPEYTDWTAIFTRDQLCGATTCDTLGMADVGMCDPKRSCSV 347
DB 337 TLKNFCKWQHQNGLDDHDEHYDAAILFTREDLCGHSHCDTLGMADVGTICSPSCAV 396
QY 348 IEDDGLSAFTAHGLHFNHNDHNVKVEEVFGKLRANHMSPPTLIQIDRANPWSACS 407
DB 397 IEDDGLHAAFTVAHEIGHLLGSHDSKFCENFGTTEDKRLMSILTSIDASKPWSKCT 456
QY 408 AAILTFDLDGSHGCLLDOPSKPISLPELPGASYTILSOCELAFYGVSKPCPYMQYCTK 467
DB 457 SATITFELDDGHCNLLDLPRKQLGPEELPGQYTDATQCNLTGPEYSCVCPGMDYCAR 516
QY 468 LWTGKAGQMVCOTRHPFADGTSCGEGKLCCLKGACVER-----HNLNKHRYDVGSWAKW 522
DB 517 LWCAVVRQGMVCLTKKLPAVEGTPCGKRGVCLQKGVCKDKTKKYYTSSH---GNWGSW 573
QY 523 DPYCPGSRCTGGGVLARRQCTNPTPANGKYCEGVRYKYSRNLNLEPCPSASGKSFREE 582
DB 574 GPWQCSRSRGGGVQFAYRHCNPNAPRNSGRYCTGKRAIYRSCSVTPCP---PNGKSPRHE 631
QY 583 QCEAFNGYNHSTRLTLAVAWPKYSGVSPRDKCLCRANGFYGVFLAPKVVVDGTLCS 642
DB 632 QCEAKNGYQSDAKGVKTFVWPKYAGVLPADVCKLTCRAKGHYVYVPSPKVTDGTECR 691
QY 643 PDSTSVCVQKCKTAKGCDNGLSKRFDKGCVGGDNKSKCKVTG 687
DB 692 PYSNSVCVRGCRVTDGDIIGSKLOYDKGCVGGDNSSCTKIIG 736

RESULT 8
ATSS_HUMAN
ID ATSS5_HUMAN STANDARD; PRT; 930 AA.
AC Q9UNA0; Q9UKP2;
DT 16-OCT-2001 (Rel. 40, Created)
DB 16-OCT-2001 (Rel. 40, Last sequence update)
```

```
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (ADAM-TS 11).
GN ADAMTS5 OR ADMP2 OR ADAMTS11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99367476; PubMed=10438522;
RA Abbaszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H.,
RA Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M.,
RA Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K.,
RA Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H.,
RA Ito Y., Newton R.C., Magolda R.B., Trzaskos J.M., Hollis G.F.,
RA Arner E.C., Burn T.C.;
RA "Cloning and characterization of ADAMTS11, an aggrecanase from the
RT ADAMTS family.";
RL J. Biol. Chem. 274:23443-23450(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Sasaki T., Shibuya K., Kawasaki K., Asakawa S.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Schaefer M., Schoen O., Desario A., Reichelt J., Kauter G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Daggand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Vaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE OF 413-930 FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.T., Hirohata S., Seldin M.F., Apte S.S.;
RA "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT zinc metalloproteinases.";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
CC PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393
CC site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
CC CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,
CC CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
CC ARTHRITIC PATIENT.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```


CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
 CC SIMILARITY).
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A PURIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF261918; AAF89106.1; -
 CC EMBL: AB037733; BAA92550.1; -
 CC HSP: P15167; IATL.
 CC MEROPS: M12.021; -
 CC Genew: HGNC:13202; ADAMTS9.
 CC MIM: 605421; -
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR002870; Pep_M12B_propep.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSPI.
 CC InterPro: IPR000130; Zn_M12peptdse.
 CC Pfam: PF00090; tsp_1; 11.
 CC Pfam: PF01421; Reprolysin; 1.
 CC Pfam: PF01562; Pep_M12B_propep; 1.
 CC SMART: SM00209; TSPI; 12.
 CC PROSITE: PS02015; ADAM_MEPRO; 1.
 CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC PROSITE: PS00092; TSPI; 9.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; zymogen;
 CC Repeat: Extracellular matrix; Alternating splicing.
 CC SIGNAL: 1 18
 CC PROPEP: 19 287
 CC CHAIN: 288 1629
 CC DOMAIN: 509 587
 CC DOMAIN: 589 642
 CC DOMAIN: 645 752
 CC DOMAIN: 753 880
 CC DOMAIN: 999 1053
 CC DOMAIN: 1056 1108
 CC DOMAIN: 1111 1156
 CC DOMAIN: 1184 1239
 CC DOMAIN: 1240 1295
 CC DOMAIN: 1332 1383
 CC DOMAIN: 1386 1439
 CC DOMAIN: 1445 1498
 CC DOMAIN: 1501 1554
 CC DOMAIN: 1562 1612
 CC DOMAIN: 88 96
 CC SITE: 223 223
 CC METAL: 434 434
 CC ACT_SITE: 435 435
 CC METAL: 438 438
 CC METAL: 444 444
 CC CARBOHYD: 112 112
 CC CARBOHYD: 135 135
 CC CARBOHYD: 271 271
 CC CARBOHYD: 749 749
 CC CARBOHYD: 840 840
 CC CARBOHYD: 1213 1213
 CC CARBOHYD: 1267 1267
 CC VARSPLIC: 1064 1072
 CC
 CC VARSPLIC: 1073 1629
 CC CONFLICT: 367 367
 CC SEQUENCE: 1629 AA; 182649 MW; C1C4CEFF58B8941F CRC64;

Query Match 41.2%; Score 1543.5; DB 1; Length 1629;
 Best Local Similarity 42.5%; Pred. No. 6e-107;
 Matches 308; Conservative 99; Mismatches 246; Indels 71; Gaps 13;

QY 22 EREVVVPIRLDP-----DINGRRYV-----RGPESDGOGGLFOITATQ 61
 DB 47 EYEIVSIRVNALGEPPTVNHFKRTRRSINATDWPAPASSSSSTSSQAHYRUSAF 106
 QY 62 EDFYHLTPDAQFIAPAFSTEHLGVP-----LOGLTGSSDLRCFYSGDVNAPDPSFAAV 117
 DB 107 QQELENLTANAGFIAPLFTVLLGTGPGVNTKYESEAEKHKCFYGVYVNTSEHTAVI 166
 QY 118 SLGGLRGAGYGAEVISPLNAPAAQRNSQAHLQRRGVPGSGDPTSCGVA 177
 DB 167 SLGGMGLTFRSHGDYFIEPLQSMDEQDEDEQNPHIYRRSAFQREPTSGRHACDTS 226
 QY 178 SGWNPAILRALDPYKPRRAGFE-----SRSSRRSG 208
 DB 227 EHKN---RHSKDKKTKRKWGERINLAGDVAALNSGLATEAFSAVGNKTDNTRERTHR 283
 QY 209 RAKRFVSIPIRYVETLVVADESVMYKFGADLEHYLLTLATAARLYRHPISILNP 268
 DB 284 RTKRFLSYPRFEVLVADNRNVMVSYHGENLQHYILTMSIVASYIKDPSIGNLINIVVN 343
 QY 269 VLLLRDRDSGPKVTGNAALTNRFCAWOKKLNKVSQKHPEYWDTAILFTRODLGCA-TTC 327
 DB 344 LIVIHNEQDQPSISFNAQTTLKNFCOMQHSKNSPGGIH---HDTAVLLTRQDTCRAHDKC 400
 QY 328 DTLMADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEVEFGKLRAN 387
 DB 401 DTGLAELGTICDPYRSCSISEDSGLSTFTIAHELGHVFNMPHDNDNNCKEE-GYKSPQ 459
 QY 388 HMSNPTLIQIDRANPWSACSAIITDLSGHGDCILLDQ-SKPISLPEDLPASVTLQ 446
 DB 460 HVMAPTILNYTNPMWMSKCSKYITEFLDTGYCECLLNPEPSRPYPLPVQLPGILYNVK 519
 QY 447 QCELAFCVSGSKPCPYQWYCTKLWC---TGKAKQWVCOTRHPFWADGTSCGEGKCLKGA 503
 DB 520 QCELIFGQGVQVPMQCRLLMCNNVNGVHKG---CRTQHTPWADGTCEPGKHKCYGF 576
 QY 504 CVERHNLNHRVDGSKWAKWDYPGCSRTCGGVQLARRQCTNTPANGKCYGCVKVKYR 563
 DB 577 CVPK-EMDVPTDGSNGSWSPFGTCSRTCGGKIKTAIRECNPENKNGKCYCVGRMKFK 635
 QY 564 SCNLEPCPSASGKSFPREOCEAFNGYNNSTNRLTLAVAWVPKYSGVSPDRCKLCFRAN 623
 DB 636 SCNTEPCLKQK--RDFRDEQCAHFDGKHFNGLLPNVRWVPKYSGLMKDRCKLCFRVA 693
 QY 624 GTGYEYVLAPKVVVDGTLGSPDSTVSVQVQKCIKAGCDGNGLSKKRDKCGVGGDNKSK 683
 DB 694 GNTAYTQLDRVDIGTGPCQDINDICVQGLCROAGCDHVLNSKARRDKCGVGGDNSSCK 753
 QY 684 KVTG 687
 DB 754 TVAG 757

RESULT 10
 ATSA_RAT
 ID ATSA_RAT STANDARD; PRT; 630 AA.
 AC Q9ESP7: Q9ESP8; Q9ESP6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-4 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
 DE (Fragment).
 GN ADAMTS4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;

RN [1] SEQUENCE FROM N.A.
 RP STRAIN-Wistar; TISSUE=Brain;
 RX MEDLINE=20415831; PubMed=10961658;
 RA Satoh K., Suzuki N., Yokota H.;
 RT "ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin motifs) is transcriptionally induced in beta-amyloid treated rat astrocytes";
 RL Neurosci. Lett. 289:177-180(2000).
 CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES.
 CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393 site.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
 CC -1- TISSUE SPECIFICITY: ONLY EXPRESSED IN BRAIN.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB042272; BAB16474.1; -;
 CC EMBL; AB042271; BAB16473.1; -;
 CC EMBL; AB042273; BAB16475.1; -;
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSPL.
 CC InterPro: IPR000130; Zn_MTPeptidse.
 CC Pfam; PF00090; tsp.1; 2.
 CC SMART; SM00209; Reprolysin; 1.
 CC PROSITE; PS0215; ADAM_MEPRO; 1.
 CC PROSITE; PS0092; TSPL; 1.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC Hydrolase; Metalloproteinase; Zinc; Glycoprotein; Zymogen;
 KW Extracellular matrix.
 FT NON_TER 1 1
 FT PROPEP <1 5 BY SIMILARITY.
 FT CHAIN 6 630 ADAMTS-4.
 FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 155 155 ZINC (CATALYTIC).
 FT METAL 158 158 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 164 164 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 233 303 DISINTEGRIN-LIKE.
 FT DOMAIN 316 367 TSP TYPE-1.
 FT DOMAIN 368 478 CYS-RICH.
 FT DOMAIN 479 630 SPACER.
 FT DOMAIN 40 45 POLY-ALA.
 FT CARBOHYD 96 96 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 474 474 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 630 AA; 68384 MW; 634428753167CFE7 CRC64;
 Query Match 40.6%; Score 1518.5; DB 1; Length 630;
 Best Local Similarity 55.8%; Pred. No. 1.4e-105;
 Matches 271; Conservative 67; Mismatches 137; Indels 11; Gaps 6;
 QY 209 RAKRFVSIPIYETLVVADESQVYFGADLEHYLLLTAAARLYRHPSPILNPINIVVVK 268
 DB 2 RTRKFAFSLRFVETLVVADKMAAFHAGLKHVLLVMAAAAKAFRHPSPINPVLVWTR 61

QY 269 VLLLRDRSGPKVTGNAAITLRFNFCAMOKLKNKYSDKHPEYWDTAILFTRODLGATCTD 328
 DB 62 LVILSGQEVPOVGPVSAATLRSFCTWQKGLNPPNDSPPDHFDTAILFTRODLGATCTD 121
 QY 329 TLGHADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVCEVFGK-LRAN 387
 DB 122 ALGMAGVGTGCDPARSCAIVEDDGLQSAFTAHELGHVFNHLDNSKPCANLNGQGSRR 181
 QY 388 HMSPTLTIQIDRANPWSACSAIITDFLDSGHGDCILDQPSKPISLPEDLPFGASYTILSQ 447
 DB 182 HVMAPVMAHVDPEEPWSPCSAREITDFLDNGYGHCLLDKPEALHLPVTFPGDYDADQ 241
 QY 448 CELAFGVGSKPCPYM-QYCTKLMTGKAKGMVQOTRHPFWADGTSGEGLKLUKACVE 506
 DB 242 QLTFTGPDSSHCPQLPPPCAAALWCFGLNHGAMQOTRHPFWADGTPCGPAQACMGRC 301
 QY 507 RHNLNKHVD--GSWAKWDVPYGCSTCGGVQVQLARQCTNPTPANGKYGCEGVRYKRS 564
 DB 302 VDQLKDFNIPQAGMGPMGDCSRCTCGGVQVFSRRDCTAPVPRNGKYGCEGRRTFERS 361
 QY 565 CNLEPCSSASGKSFREPCQCEAFNGYNHSTNRLTL---AVAWPKYSGVSPDRCKLICR 621
 DB 362 CNTKNCP-HGSALTFREPCAA---YNHRTDLKSFPGPMQWVRYTGVAPRQCKLTQ 417
 QY 622 ANGTYFYVLAPKVVDGTLCSFDSSTVSCVQCKICAKGCDNGLSKKRFKCGVCGDNKS 681
 DB 418 ARALGYVYVLEPRVADGTPCSPDSSVSCVQVGRCHAGCDRIIGSKKRFKDMCGGNGSS 477
 QY 682 CKKVTG 687
 DB 478 CSKQSG 483
 RESULT 11
 AT57_HUMAN
 ID AT57_HUMAN STANDARD; PRT; 997 AA.
 AC O9UKP4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
 GN ADAMTS7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93935124; PubMed=10464288;
 RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
 RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of Zinc Metalloproteinases";
 RL J. Biol. Chem. 274:25555-25563(1999).
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

```

or send an email to license@isb-sib.ch).
-----
CC CC EMBL; AF140675; AAD56358.1; -
DR HSSP; P15167; 1ATL.
DR MEROPS; M12.231; -
DR MIM; 605009; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptdse.
DR Pfam; PF00090; tsp_1; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPl; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00215; ADAM_MEPHO; 1.
DR PROSITE; PS00092; TSPl; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 BY SIMILARITY.
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match 29.08; Score 1087; DB 1; Length 997;
Best Local Similarity 38.68; Pred. No. 3.2e-73;
Matches 286; Conservative 82; Mismatches 264; Indels 108; Gaps 32;

QY 1 MLLGILFLAFAGTAGGFEPEP---EWWPIRLDP-----DINGRR- 39
D 18 LLLLCALAPAGPAPGATGTEGRAALDIVHPVRVDAGSFLSYELWFLAKRKRDVYRDR 77
QY 40 ---YY---WRGPESDGOGLIFQITAFQEDPYLHLTPDAQFLAPAF--STEHLGVPLQGL 91
D 78 APAFYELQYRGRE-----LRFNLTAQ-----HL-----LAPGFVSETRRRG---GL 116
QY 92 TGGSSDLR---RCFYSGDVNAEPD---SFAAVSLCGGLRGAFYRGAEYVISPLNPA 144
D 117 --GRAHIAHTPACHILGEVQ--DPELGGLAAISACDGLKGVFOLSNEIDFIEPLD--SA 171
QY 145 PAAORNSOGAHLQRRGV-----GGPSGDPTRSGVAGWNPAIRALRDPYKPRRAGFG 199
D 172 PAREPQAQ--PHVYKQQAPELAQRGDSSAP--STCGV-----QVYPELESRRERW 219
QY 200 ESRRRRSRGRKRF---VSPRPRVETVLVADESQWVFGH--ADLEHYLLLLTAARLYR 254
D 220 EQQRQRPRRULHQRSVSKERKWTETLVADAKMVIYHGQPVQVESYVLTIMNVAGLFH 279
QY 255 HPSILNFINIVVKKVLLLRDRSDPKVTGNAALTFLRNCFAWKLNKVSQKHPEYWDAT 314
D 280 DPSIGNPHITIVRLVLEDEEDLKITHADNTLKSFCWKQSKINMKGAHPLHDDTAT 339
QY 315 LFTQDLCGATT---CDTLGMADVTGMDCKRSCSVIEDDGLPSAFTTAHHLGHVFNMPHD 372
D 340 LLTRKDLCAAMNRPCTETLGLSHVAGMCQPHRSCSINEDTGLPLAFTVAHELGHLSFGIQHD 399
QY 373 -NVKVEEVFGKLRANHHMSTPLLIQDRANPWSACSAAITDFLDSHGDCLLDQPSKP- 430

```

```

Db 400 GSGNDCBPV-GK--RPFIMPSPQLLYDAAPTWSRCSROYITRFLDRGWGLCLDDPPAKDI 456
QY 431 ISLPEDLPGASYTLTSLQOCELAFYGVGSKPCPYM-QYCTKLMCTGKAKGQWVCQTRHFPWAD 489
D 457 IDPFSVPPGVLYDVSHQCRLOYGAYSAFCEMDMNVCHTLWCSVGT-----TCHSKLDAVD 512
QY 490 GTSCGEGKCLKGACVERHNLNKHVRDGSNAKWDPYGPCSRTCGGGVOLARRQCTNPTPA 549
D 513 GTRCGENKWLCSGECVP-VGFRPEAVDVGWSGWSANSICSRSCGCMGVQSAERQCTQTPPK 571
QY 550 NGKGYCRGVKVRKSCNLEPCSSASCKSPREEQCEAFNGYNHSTNRLTLTAVAVWPKYSG 609
D 572 YKGRYCVGERKFRFLCNLQACP--AGRPSFRHVOCSEHFDAMLYKGQLHT-----WVPVND 635
QY 610 VSPRDKCKLICRANGTGYFYVLAPKVVVDGTLCSPP--DSTSVQVQGVKICAKCGDNLGSKK 667
D 626 VNP---CELHCRPANEFYFAKKLRDVAVDGTPCYQVQRASRDLCINGICKNVGCDFEIDSGA 682
QY 668 RFDKCGVCGGDNKSCKKVTG 687
D 683 MEDRCGVCHGNGSTCHTVSG 702

RESULT 12
AT10_HUMAN STANDARD; PRT; 1077 AA.
ID AT10_HUMAN
AC Q9H324;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT thrombospondin type I repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF163762; AAG35563.1; -
DR MEROPS; M12.235; -
DR Genew; HGNC:13201; ADAMTS10.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptdse.
DR Pfam; PF00090; tsp_1; 5.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.

```

```
DR SMART: SM00209; TSPL: 5;
DR PROSITE: PS00215; ADAM_MERO; 1.
DR PROSITE: PS00092; TSPL: 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
FT Repeat; Extracellular matrix.
FT NON_TER 1
FT PROPEP 1
FT CHAIN 1
FT METAL 208 1077
FT ACT_SITE 366 366
FT METAL 366 366
FT METAL 370 367
FT METAL 376 376
FT METAL 376 376
FT METAL 434 520
FT METAL 578 679
FT METAL 580 802
FT METAL 521 577
FT METAL 799 860
FT METAL 862 918
FT METAL 922 976
FT METAL 981 1031
FT METAL 64 64
FT METAL 196 196
FT METAL 297 297
FT METAL 714 714
FT METAL 769 769
FT METAL 866 866
FT METAL 1077 1077
FT METAL 28 68; Score 1072.5; DB 1; Length 1077;
FT METAL 36.88; Pred. No. 4.2e-72;
FT METAL 260; Conserved; 89; Mismatches 280; Indels 77; Gaps 23;
Query Match
Best Local Similarity
Matches 260; Conserved; 89; Mismatches 280; Indels 77; Gaps 23;
QY 24 EVVPIRLDPDINGR-----RYWGPEDSGDGLFIQITAFQEDFYHLHTDAQF 74
DB 14 EIAFPTVDH--NGALLAFSPPPRRRGRTGATAESLFYKVASPTHELLNLTSSRL 71
QY 75 LAPASTHGLVPGQLTGGSSDLRCFYSGDVNAE--PDSFAVSLCGGLRGAFYRGAE 133
DB 72 LAGHVSVEYH--TREGLAQRAARPHCLYAGHLOGQSSSHVAISTCGGLHGLIVADEEE 129
QY 134 YVISPPLNAPAAORNQGAHLQRRGVPGSGDPTSRGV--ASGW--NFAILRALD 189
DB 130 YLIEPLHGKPGSRSPESGPHVYVYKSLRLHPLD--TAGGVYDERPKWGRFWLRTLK 187
QY 190 PKPRPAGFGSRSSRRSGRAKRVFIPRYVEILVWADSWKPHG--ADLEHYLLFLAT 248
DB 188 P--PPARPLGNETERGQPG--LKRVSRSRYVETLVADKMWVYHGRDRVQYVLAIMNI 244
QY 249 AARLYRHPISILNPINIVVVKVLLLRDSDGPKVTGNALTLRNFCAWQKL-----NK 301
DB 245 VAKLFQDSLSGSTVNILYTRILITLITEOQPTLEITHHACKSLDSFCKWQKSVNHSHGNA 304
QY 302 VSKHPEYWDYAILFTRODLC--GATCTDILGMADVTCMDPKRSCSVIEDDGLPSAFT 359
DB 305 IPGVANHHDTAVLITRYDICIYKKNKPGCTGLAPVGMGERSCSVNEDIGLQPAFTI 364
QY 360 AHELGHVFNMPHDNVKVEEYFG---KLRANHMMSPTLIQIDRNP--WSACSAALITD 413
DB 365 AHEIGHTFGMHDGVNCGARGQDPKALMAAHITMTK-----NPFVWSSCNRYIYS 417
QY 414 FLDSGHLGLLDQPSK--PISLPEDLPGASYTLSSQCLAFGSKPCPYMQYCTKLWCTG 472
DB 418 FLDSGLGLCLNRRPPQDFVYPTVAPGQAYDADQCRFHQGVKSRQCKYGEVCSLWCLS 477
QY 473 KAKQWQVCTHFFWADTSCGEKL-----CLKACACVERINLKHVRDGSNAKWIDYGPC 528
DB 478 KSNR---CITNISIPAAEETLCQTHIDKGMWCYKRVCP--EGSRPEGVDGAWGFWTPWDC 533
QY 529 SRTCGGGVQLARRQCTNPTPANGKYGEGVVKVRSNCLPCPSSAGSKGFRFEEQCEAFN 588
DB 534 SRTCGGGVSSSSRRHCDSPRTIGKCYCLGERRRHSNTDDCPGS--QDFREVQVSEFD 591
```

```
QY 589 G-----YNHSTNRLTLAVAWVPKYSVSPRDKCKLICRANGTGYFYVLAPKVDGTLC 641
DB 592 SIPFRGKFKWKTYR-----GGVK---ACSLTSLAEGNFYTERAAAVVDGTFC 638
QY 642 SPDSVSCVQCKICIKAGCDNLGSKKRFKFDKCGVCGGDNKSKCKVTG 687
DB 639 RPDVTDICVSGECKHVGCDRVLSGLREDKRCVCGGDSACETIEG 684
RESULT 13
AT12_HUMAN
ID AT12_HUMAN STANDARD; PRT; 1593 AA.
AC P58397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S. Argueelles J M, Fernandez P L, Lopez-Otin C.;
RT "Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).
CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely expressed in gastric carcinomas and in cancer cells of diverse origin.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -!- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1 DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
EMBL; AJ250725; CAC20419.1;
Genew; HGNC:14605; ADAMTS12.
MI; 606184;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPL.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF00090; tsp_1; 6.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPL; 8.
```


Result No.	Query	Score	Match	Query			ID	Description
				Length	DB			
1	3736	99.8	950	4	Q8TE58	Q8TE58	homo sapien	
2	1223	32.7	2165	5	Q19791	Q87931	caenorhabditis	
3	1022	27.3	1688	5	O8SXB0	Q8xb01	drosophila	
4	1015.5	27.1	1081	4	Q8TE60	Q8te60	homo sapien	
5	1015	27.1	1054	5	Q9W493	Q9w493	drosophila	
6	986.5	26.7	1095	4	Q8TE56	Q8te56	homo sapien	
7	989.5	26.4	1229	5	Q9VF61	Q9vfe1	drosophila	
8	967.5	25.8	1072	4	Q8TE57	Q8te57	homo sapien	
9	941.5	25.1	1207	4	Q8TE59	Q8te59	homo sapien	
10	928	24.8	1223	4	O8WXS8	Q8wx38	homo sapien	
11	928	24.8	1223	4	Q8TE55	Q8te55	homo sapien	
12	926.5	24.7	1159	4	Q8TEY8	Q8tey8	homo sapien	
13	830	22.2	169	6	Q9GL54	Q9gl54	oryctolagus	
14	720	19.2	1427	4	Q9GL37	Q9gl37	homo sapien	
15	558	14.9	192	6	Q95N24	Q95n24	equus caball	
16	518	13.8	187	6	Q95N23	Q95n23	equus caball	

```
|||||
Db 121 GGLRGAFYGAAYVISPNSAPAAQNSOGAHLQRRGVPGPSGDPTSCGVASGW 180
QY 181 NPAILRALDPYKPRRAGFESRRSRGAKRFVSPRYVETLVVADESVMKFGADLEH 240
Db 181 NPAILRALDPYKPRRAGFESRRSRGAKRFVSPRYVETLVVADESVMKFGADLEH 240
QY 241 YLTLTATAARLYRHPSILNPINVVVYLLLRDRSDGPKVTGNAALTIRNFCAMQKLN 300
Db 241 YLTLTATAARLYRHPSILNPINVVVYLLLRDRSDGPKVTGNAALTIRNFCAMQKLN 300
QY 301 KVSDEKHEPYWDTAITRQDLGATTCDTLGMADVGTMDCKPRSCSVIEDDGLPSAFTTA 360
Db 301 KVSDEKHEPYWDTAITRQDLGATTCDTLGMADVGTMDCKPRSCSVIEDDGLPSAFTTA 360
QY 361 HELGHVFNPHDNNVKEEVEFGKLRANHMSPTLQIDRANPWSACSAIIIDFLDSHG 420
Db 361 HELGHVFNPHDNNVKEEVEFGKLRANHMSPTLQIDRANPWSACSAIIIDFLDSHG 420
QY 421 DCLLDQSPKISLPEDLPASVYLSOCELAFCGVGSKPCPYMQYCTKLWCTGKAKQWVC 480
Db 421 DCLLDQSPKISLPEDLPASVYLSOCELAFCGVGSKPCPYMQYCTKLWCTGKAKQWVC 480
QY 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHVRDGSWAKWDPYGPCSRTCGGVQLAR 540
Db 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHVRDGSWAKWDPYGPCSRTCGGVQLAR 540
QY 541 RQCTNTPPANGKYCEGVVRYKYSNLEPCPSSASGSKSFREEOCEAFNGYNHSTNRLTLA 600
Db 541 RQCTNTPPANGKYCEGVVRYKYSNLEPCPSSASGSKSFREEOCEAFNGYNHSTNRLTLA 600
QY 601 VAWPKYSGVSPDRCKKLICRANGTYFYVYLAAPKVVVDGTLCSPDSTSVCVQKCIKAGCD 660
Db 601 VAWPKYSGVSPDRCKKLICRANGTYFYVYLAAPKVVVDGTLCSPDSTSVCVQKCIKAGCD 660
QY 661 GNLSKRRFDKCGVCGGDNKSCCKVTG 687
Db 661 GNLSKRRFDKCGVCGGDNKSCCKVTG 687

RESULT 2
Q19791 PRELIMINARY; PRT; 2165 AA.
ID Q19791; Q27524;
AC Q19791-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gajadaty S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
EX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
```

```
EMBL; Z69361; CAA93288.1; -.
EMBL; Z69360; CAA93288.1; JOINED.
EMBL; Z69360; CAA93287.1; -.
EMBL; Z69361; CAA93287.1; JOINED.
HSP; P15167; IOTH.
MEROPS; M12.135; -.
DR InterPro; IPF001590; Reprolysin.
DR InterPro; IPF000884; TSPI.
DR InterPro; IPF00130; Zn_Mtpeptidse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00080; tsd_1; 14.
DR SMART; SM00209; TSPI; 18.
DR PROSITE; PS00215; ADAM_MEPPO; 1.
DR PROSITE; PS00092; TSPI; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;

Query Match 32.78; Score 1223; DB 5; Length 2165;
Best Local Similarity 40.5%; Pred. No. 1.4e-92;
Matches 257; Conservative 87; Mismatches 214; Indels 76; Gaps 16;

QY 101 CFYSGDVNAEPDSFAAASLC---GGLRGAFYGAAYVISPNSASA---PAAQNSOGA 154
Db 169 CIYFAHVKG-VHSHIVNLCDSEDLGYMLALPSGIHTVEPIISGNGTEHDGASRHQ-- 225
QY 155 HLLQRRGVPGPSGDPTSCGV-----ASGNPAILLRALDPYKPRRAGFESRSR 204
Db 226 HLVRKEDPMHFKSFDFHLSNSTSVNETTETVATWQDQWEDVIER-----KARSR 272
QY 205 RRSGRAKRFVSPRYVETLVVADESVMKFGADLEHLLTLATAARLYRHPSILNPINI 264
Db 273 RAAN-----SWDHVEVIVVADTKMYEYHGRSLEDYVTLFSTVASIYRHSRASINV 326
QY 265 VVVKVLLLRDRSDGPKVTGNAALTIRNFCAMQKLNKVSDEKHEPYWDTAITRQDLGCA 324
Db 327 VVVKLVLTENAGPRITQNAQOTLQDFCRWQYYNDPDDSVQHHDAIILLTRKDIORS 386
QY 325 T-TCTLTGMADVGTMDCKPSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVKEEVEV--- 380
Db 387 QGKCDTLGLAELGTMDCKMOKSCAIIEDNGLSAAFTIAHELGHVFSIPHDDERKCSYMPV 446
QY 381 -----FGKLRAN---HMSPTLIQIDRANPWSACSAIIIDFLDSHG--DCLLD 425
Db 447 NKVCKFQSTKFDKTOFQNNFHIMAPTLEVTNHPWSWSPSCSAGMLERFLENNRGOTQCLFD 506
QY 426 QPSKPSILPE-----DLPGASYTLSSOCELAFCGVGSKPCPYMQYCTKLWCTGKAKQWVCQ 481
Db 507 QPVERRYVEDVFRDEPGKKYDAHQCKFVFGPASELCFPMPTCRLMLCATFYGSQMGR 566
QY 482 TRHFPWADGTSCEGK--LCLKGACVERHNLNKHVRDGSWAKWDPYGPCSRTCGGVQLA 539
Db 567 TQHPWADGTPCDESRMFCFHGACVRLAPESLTKIDGQWDMRSWGECRSRTCGGVQKG 626
QY 540 RRQCTNTPPANGKYCEGVVRYKYSNLEPCPSSASGSKSFREEOCEAFNGYN-----HS 593
Db 627 LRDCDSPKPRNGKYCVGQREYRSCNTQECPWDT--OPREVQCSSEFNKNDIGIQGVAS 684
QY 594 TNRLTAVAWPKYSGVSPDRCKKLICRANGTYFYVYLAAPKVVVDGTLCSPDSTSVCVQK 653
Db 685 TN-----THWPKYANVAPNERCKLYCRUSGSAAFVLLLRKVVDGTPCDNRGDDICVAGA 739
QY 654 CIKAGDGNLGSKKRFDKCGVCGGDNKSKCKVTG 687
Db 740 CWPAGCDHQLHSTLRDCKGVCVGGDDSSCKVKVG 773

RESULT 3
Q8SXB0 PRELIMINARY; PRT; 1688 AA.
ID Q8SXB0;
AC Q8SXB0;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
```

DE GH16393p.
GN CG6107.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCHI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY094716; AAMI1069.1; -
SQ SEQUENCE 1688 AA; 198867 MW; 48FB8DD4DE0CA4D2 CRG64;

Query Match 27.3%; Score 1022; DB 5; Length 1688;
Best Local Similarity 35.1%; Pred. No. 5.8e-76;
Matches 237; Conservative 104; Mismatches 237; Indels 98; Gaps 24;

QY	55	FQITAFQEDFYLHLTPDQAFAPASTSHLGVPLQG-----LTCGSSDLRR--CFYSGD	106
Db	296	YNLNVFGRLHLVLRQDAFV-HNHSMTTHILKEGEGHPGEPETAEEAQRHLGCFYSGY	354
QY	107	VNAEPDSPAAYSLCGGLRGATGYRGAEVVIGPLPNASAP-----AAORNSOGAHLIQ	158
Db	355	VEDDPHSMVSVSLCGMGTYIKTSGALLIQPVNRTSDEVLHVRFRKSQRNAR--HAVS	412
QY	159	ARGVPGSPSGDPTSRGCVASGNPAILRALPDYKPRRAGFG-----SRSRRSRGRKRFV	214
Db	413	KFEL-----GLDFMSKLEQVQEEQKSKSLNRKKRHYA	448
QY	215	SIPRTV---ETLVVADESVMVKGADLEHYLLTLATAARLYRHPISILNPVVVKVL	271
Db	449	DVDNQVYTVLEVIADVNSMKQFHGEDLPQYILILMSIVSSIFADAGISNIRILVRLIS	508
QY	272	LRDRDSGPKV---TGNAAALTURNCAMQKLNKYSDKHPYWDTAILETRDQLCGAT---	325
Db	509	L-----PNINDQTHSSNEMLKHFQF---INQSGYR---DTAMLITREPCGSVPK	555
QY	326	TCDTLGMADVTCMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPNHDNWKVCEEVFGKL	385
Db	556	ICHMLGLAELGTVC--SSCSIVDQTLGPTAFTWAHELGHILNNMHHDDDKCMPTVTRQN	614
QY	386	AN---HMMSPTLIQIDRANPWSACSAALIITDLSGHGCDLDDQPSKIPSLP---EDLPG	439
Db	615	NNKVLHIMSSVMGIMHPWWSKCSRHPVSFELEKTDKSL--ETSVCAHIPTYGTERLPG	672
QY	440	ASYTLSQOCELAFGVSGPKCPYMQYTKLWC---TGKAKGMVCQTRHFPWADGTSCG-E	495
Db	673	EIYSLDAQCLSFNGDFGTCPTDECKRLWCNRTSGNSNEQ--CASSNLPWADGTPCGSS	730
QY	496	GKLCIKGACVERHNLNKH---RYDGSNAKWDPYPCSRCTCGGVQLARROCTNPTPANG	551
Db	731	GHWCQRKCVS---NKHGCGRQVNGGWGPTPTPCSLTCGGGVQESRECNQDPVPENG	786
QY	552	GKYCEGVKVKYRSCNLEPCPSASGKSPREOCEAFNGYNSHTRLTTLAVAWVPKYSVGS	611
Db	787	GKYCTGSKKYRSCNTHQCPGSMDP--REOQCVMNGRNMNI PGVNPDTKWVPKYE---841	
QY	612	PRDKCKLICRANGTYFVFLAPKVVDGTLCSFDSTSVCVQKGCIKAGCDGNLGSKKRFDK	671
Db	842	-KDACILFCRMDMKVTFYMLKSMYTDGTSCAVDSFDKCVNGICRPAGCDNELNSTAKLDK	900
QY	672	CGVCGGDNKSKCKVTG	687
Db	901	CGVCEGRNDTCHETVG	916

```
QY 596 RLTLAVAWPKYSGVSPRDKKLCICRANGTGYFYVLAPKVVDTGLCSPDSTSVCVQGHKI 655
Db 669 -----WKP-VTKVEEDRCKLYCKAENFEFFFAKGVKVDGTPCSPKNDVCDIGVCE 720
QY 656 KAGCDNGLSKRFKCKGCGVGDNKSCKKVTG 687
Db 721 LVGCDHELKSKAVSDACGVCKGDNSTCKFYKG 752

RESULT 5
Q9W493 ID Q9W493 PRELIMINARY; PRT: 1054 AA.
AC Q9W493;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE CG4096 protein.
GN CG4096.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortenson J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers J.R., Blazef R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaesner K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Rees M.G.,
RA Reiner T.C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003435; AAF46065.1; -
DR HSP; P15167; IATL.
DR MEROPS; M12.231; -.
DR FlyBase; FBgn029791; CG4096.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPL.
DR InterPro; IPR000130; Zn_Mtpeptdase.
```

```
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF00090; tsp_1; 2.
DR SMART; SM00209; TSPL; 2.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPL; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN.1.
SQ SEQUENCE 1054 AA; 118616 MW; DC15455555CB6212 CRC64;

Query Match 27.1%; Score 1015; DB 5; Length 1054;
Best Local Similarity 33.2%; Pred. No. 1.2e-75;
Matches 266; Conservative 93; Mismatches 269; Indels 174; Gaps 29;

QY 17 GGFE-----EREVVPIRLDPD-----INGRRYVVR-----GPED 47
Db 23 GGVRLYGLHSDLVAGEQLVPRVHPDGAFTWOLEYAHLDHRRHRQRSLNSEHD 82
QY 48 SGDGLIFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRCFFSGDV 107
Db 83 TQADLHLHLLPLANETHLELMAHSYFLAPNLVVERHRRDLRTSRPLTTRHLNCHPHGV 142
QY 108 NAEFDSFANVSLCGGL-----RGAFYGAEVVISPL-----PNASA 144
Db 143 RGQFATNVAISTCAGLVSCPFCVLPELITLCQVGHIRTAGNEYFIEPSEHEHPHNVGH 202
QY 145 P--AAQRNS-OGAHLQRRG--VFGPSGD--PTSRCGVASGMNPAILRALDPYKPRR-- 195
Db 203 PHVFORSVKPKHSLRKNRKRKGGKSGSGAEVSNCGTR-----EPRRM 248
QY 196 -----AGGES 201
Db 249 ETRLEWQARGKVKVQGGQIRRHRRHHHHHHKHYRHHOOKISRVPHTKFYETQFTE 308
QY 202 RSRRSRGAKRFVSTPRVETLVVADESMKVFHGADEHYLLTLATAARLYRPSILNP 261
Db 309 PDHAIEPRRRRSISSPRHETLIVADATMSAFH-RDLNGYLLTLMNVMVSLYKDPSTGNS 367
QY 262 INIVVVKVLLLRDRDSGPK--VTGNAALTLRNFCAWQKLNKVDKHPYEDTALFTRQ 319
Db 368 IEIVVRIIQLDEESQQLNLNTQNAKNLDRFCSWQHLKNGSEKDPHHHDVAILITRK 427
QY 320 DLCGATTCDTLGMADVGTMCDPKSCSVIEDGLPSAFTHLGHGVFMNPHDNVQV-CE 378
Db 428 NTC-ANNCWTGLANVGMCKPKQSCSVNEDNGIMLSHTITHLGHNFGMFHTAKIGCH 486
QY 379 EVFGKLRANHMSPTL----IQRANPWSACSAIITDFLDGSGDCLLDQPSKPI--- 431
Db 487 PRVGPPI--VHIMTPFGADTLQV-----CWSNCSRKYITHFDDQGLGEC-LDDPTPLDEY 539
QY 432 SLPELDLPGASYTLQQCELAFG-----VGSKPCPYMOYCTKLCTKAKGQMVQCTRHF 485
Db 540 NTGELPGHRYNARGCQRLQFNLTDTSEVGACSAFH-EFCSTLWC--KVNGE--CVTHMR 594
QY 486 PWADGTSCGEGKLCILKAGACVERHNKLRVDSGWSAKWDPYGPCSTCGGVQLARRQCTN 545
Db 595 PTAPGTLGCRNKWCQKGVRRREL--AAVNGWGDWSEWSECSRSCGGVSTQOQRECDN 652
QY 546 PTPANGKCYCEGVRYKRYSCNLEPCPSSASGSKSFEEQCEAFN--GYNHSTNRLTLAVAW 603
Db 653 PVPANGVCEIKERKYKICRKRPCP--AEPFSFRAQOCAREDNVSYOGATYK-----W 704
QY 604 VPKYSGVSPRDKKLCICRANGTGYFYVLAPKVVDTGLCSPDSTSVCVQGHKICAGCDNL 663
Db 705 LPFFDKNNP--CKLFCSDVDDTIANWGATVLDGTPCTGLTGNNNCIDGICKKVGCDWIV 761
QY 664 GSKRFDKCGVCGDNKSCKV 685
Db 762 DSEVQDRDCGVCGSGDGOCPV 783

RESULT 6
Q8TE56 ID Q8TE56 PRELIMINARY; PRT: 1095 AA.
```

```
AC Q8TE56;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Metalloprotease disintegrin 17, with thrombospondin domains.
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains."
RL Gene 283:49-62(2002).
DR EMBL; AJ315735; CAC86016.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1095 AA; 121099 MW; A5C6540484754D5D CRC64;

Query Match 26.7%; Score 998.5; DB 4; Length 1095;
Best Local Similarity 34.3%; Pred. No. 2.9e-74;
Matches 253; Conservative 106; Mismatches 285; Indels 93; Gaps 26;

QY 2 LLLGILTLAF-----AGRTAGGPEPEREVVPIRLDPDI-----NRRRYWRGPE 46
DB 10 LVPLVLLLVWGLDPTGVGDAADVEVLPWRVRPDDVHLPPLPAAPGPRRRRPTTP 69
QY 47 DS-----GDQGLIFQITAFQEDYFHLHTPDQAFLAPAFSTEHLGVPLGLTGGSSDLRRC 101
DB 70 AAPRARPGERALLHLPAFGRLYLQLRRDLRFLSRGSEVEEAGARR--RGRPAEL--C 125
QY 102 FYSGDVNAEPDSFAVSLCGGLRGAFG--YGAEVV--ISPLPNASAPAAQNSOGAHLQL 158
DB 126 FYSGRVLGHGSLVLSACGAAGGLVGLIQGQEVQLIPLNNSQGFSGRE----HLIR 181
QY 159 RR-GVPGPGSGD---PTSRCGVASGWNPAILRALDPYKPRAGFGE-SRSRRRSGRAKRF 213
DB 182 RWSLTPSPSAEAPQPEQLCKVLT-----EKKKPTWGRSDWRERNAILR 228
QY 214 VSIPIRYVELLVADSMVKFGAD--LEHYLTLTATARLVHRPSILNIPINIVVKVLL 272
DB 229 TS-EHTVTLVVADADMYQYHGAEAQRFILTMVMVYMFQHSGLGKIKINQVTKLVLL 287
QY 273 RDRDSGPVKTGNAALTNRNFCWQK-----LNKV--SDKHPEWDTAILFTRODL 321
DB 288 RQPAKLSIGHGERSLESFCHWQNEEYGGARYLGNQVPGCKDDPLVDRAVFTRTDF 347
QY 322 C--GATTCDTLGMADVGTCDPKRSCSVIEDGLPSAFTTAHELGHVFNPMHDNVKVEE 379
DB 348 CVHKDEPCTVGIAYLGVCSAKRKCVCVLAEDNGLNLAFTIAHELGNLGNHD----DD 402
QY 380 VFGKLRANMNSPTLIQIDRANP---WSACSAIIITDFLDSHGDCLL---DQSPKPS 432
DB 403 HSSACRSHMSGEWVK--GRNPDSLSWSSCRDLENFLSKSVSTCLLTVDPRSQRTVR 460
QY 433 LPEDLPGLASYTLISQCEALFAGVGSKPCPYMAY--CTKLWCTGKAGKQWCVOTREHPWADG 490
DB 461 LPKPLGPMHYSANECQILFGNATFCRNMEHLMCAGLWCL--VEGTSCKTKLDPLDGL 518
QY 491 TSCGKGLCLKGACVERNLNKHVRDVGSAKWDPYGPSPCGGGVQLARRQCTNPTPAN 550
DB 519 TECGADKWCRAECVSKTPIPEH--VDGDSWPWGAWSMCSRTCGTGARFRKCDNPPPGP 577
QY 551 GKGKGVYRVKRNLEPCSPSSAGSKSFREOCEAFNGYHNSTNRILTAVAVPKYSGV 610
DB 578 GGTCHPGASVHAVCNELPCPKGL--PFSRQDQQAHDRLSPKKGKLLTAV-----V 627
QY 611 SPRDKCKLICRANGTGYFVILAPKVVVDGTLCSPDSTSVCVQKCIKAGCDNGLGSKKRF 670
DB 670
```

```
DB 628 VDDPCPELYCSPLGKESPLLAVDRVLDTGTCGPYETDLCVHGKCKQKIGDGIISAAKED 687
QY 671 KCGVCGDGNKSKKVTG 687
DB 688 RCGVCSGDGKTCHLVKG 704

RESULT 7
Q9VF61
ID Q9VF61 PRELIMINARY; PRT; 1229 AA.
AC Q9VF61;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CG6107 protein.
GN CG6107.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=BERKELEY;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.I., Benton P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosløk A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AF003709; AAF55199.1; -.
DR FlyBase; FBgn0038340; CG6107.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPL.
DR InterPro; IPR000130; zn_MTpeptdse.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp.1; 2.
DR SMART; SM00209; TSPL; 3.
```


GN ADAMTS19
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V., Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of seven novel human ADAMTSs, a family of metalloproteinases with disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL: A311904; CAC84565.1; -
SQ SEQUENCE 1207 AA; 134061 MW; AF36F6BF5886FDE2 CRC64;

Query Match 25.1%; Score 941.5; DB 4; Length 1207;
Best Local Similarity 32.1%; Pred. No. 1.9e-69;
Matches 247; Conservative 100; Mismatches 278; Indels 141; Gaps 28;

QY 16 AGFEPREVVPIRLDPDINGRR-----
Db 76 AAGSREVRVAPVPLEEPVEGRSESLRPPPEGEDEELESQELPRGSSGAALSPG 135
QY 40 --YWRGP-----EDSGDGLIFQITAFQEDFVHLTPDAQFLAPAFSTE 82
Db 136 APASWQPPPPPPPPPPAQAHAEPDGE-VLLRIPAFSRDLYLLLRDRGFLAPRAVE 194
QY 83 HLGVPQLGLTGGSDLR-----RCFYSGDVNAEPDFAAASVLCGLRGARGYGAEV 135
Db 195 QRNPGGPGTGAASQPPAPPAGCFYGAHLRHGPGSLASTCGGGLMGFIQLNEDFI 254
QY 136 -ISPLNASAPAA-----QRNSQGAHLQRRGVPGGSDPTSRGCVASGNPAILRA 187
Db 255 FIEPLNDTMAITGHPHRVYRQKRSMEEKVTEK-----SALHSHYCGIIS----- 298
QY 188 LDPKPRRAGGSRSSRRSGRAKRF-VSIPR--YVETLVVADESVMKFGAD-LEHYLL 243
Db 299 -DGRPA-----SRKIASGGRKRYSKYLPQYNIETVVVADPAMVSYHGADAARRFTL 351
QY 244 TLATAARLYRHSILNPINVVVYVLLLRDRDSDGPKVTGNAALTNRNFCAMQ-KKLNV 302
Db 352 TILMVFNLFQHSGLGVQVNLRIKLLIHETPELPIYIGHGKMLESECKQHFEFGK 411
QY 303 SDKHPEY--W-----DTAILFTQDLC--GATTCDFLGADVGTMDCKRSCSVIED 350
Db 412 NDHLEMTNWGEDMTSVDAAILITRKDFCVHKDEPCDVTGAYLSGMCSEKRCIIABD 471
QY 351 DGLPSAFTTAHELGHVFNPHDN-VKVCDEVFGKLRANHMSPFTLIQIDRAN--PWSACS 407
Db 472 NGLNLAFTTAHEMGNHNGHNDNDHPSCAD-----GLHMSGEWIKQNLDGVSWSRCS 525
QY 408 AAIITDLSGHDGDCILD---QPSKPISLPEDLPGASYTLSQOCELAFGVGSKPCPYMOY 464
Db 526 KEDLERFLRSKASNCLLQTNPQSVNSVMYPSKLPGMTYTADEQCQLFGLASFQCEMOH 585
QY 465 --CTKLWCTGKAGQMVQOTRHFPHWADGTSCEGKLCCLKGACVHRNLNKHVRDGSWAKW 522
Db 586 VICTGLWC--KVEGEKERTKLDPPMDGTDCLDGKWKAGECTSRSAPEH-LAGEWSLN 642
QY 523 DPGPCSRCTCGGVQLARQCNTPPANGKCYCEGVVYKRCNLEPCPSSASGSFREFE 582
Db 643 ---SPCSRTCSAGISSRERK---PGLDSEARDCNCPKQYRICENPPCPAGLPG--FRDW 695
QY 583 QCEAFNGYHNSNRLTLAVAWPKYSGVSPDK-CKLICRANGTYFYVYLAPEKVVDTLC 641
Db 696 QCOAYSVRTSSPKHIL-----QWQAVLDEKPCALFCSPVGKEQPIILLSEKVMGDTSC 748
QY 642 SPDSTSVVOGKICIRAGCDNGLSKRKREDKCGVCGDNKSKKVTG 687
Db 749 GYGLDICANGRCQKVGCDLGLSLAREDHCGVCGNGKSKKIIG 794

RESULT 10
Q8WX38
ID Q8WX38 PRELIMINARY; PRT; 1223 AA.
AC Q8WX38;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE A disintegrin-like and metalloprotease with thrombospondin type 1 motif 14 precursor.
GN ADAMTS14
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638061; PubMed=11779638;
RA Bolz H., Ramirez A., von Brederlow B., Kubisch C.;
RT "Characterization of ADAMTS14, a novel member of the ADAMTS metalloproteinase family.";
RL Biochim. Biophys. Acta 1522:221-225(2001).
DR EMBL: AF358666; AAL40229.1; -
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSPL.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF00090; Tsp_1; 4.
DR SMART: SM00209; TSPL; 4.
DR PROSITE: PS0215; ADAM_MEPRO; 1.
DR PROSITE: PS0092; TSPL; 1.
KW Signal; Integrin; Protease; Metalloprotease.
FT SIGNAL 1 22
SQ SEQUENCE 1223 AA; 133871 MW; 36394AC4D92F170F CRC64;

Query Match 24.8%; Score 928; DB 4; Length 1223;
Best Local Similarity 33.1%; Pred. No. 2.5e-68;
Matches 222; Conservative 95; Mismatches 276; Indels 78; Gaps 22;

QY 45 PEDSGDGLIFQITAFQEDFVHLTPDAQFLAPAFSTEH-----LGVPLQLTGGSSD 97
Db 98 PGRVGRHSLYFNVTVFGKELHLRLNPNRVVPGSSVEWQEDFRELFRQPLR----- 149
QY 98 LRRCFYSGDVNAEPDFAAASVLCGLRGARGYGAEVYISPLPNASAPAAQRNSQG-AHL 156
Db 150 -OECVYTGVTGMPGAAVAISNCDGLAGLRTDSTDFIEPLERQ---QEKEASGRTHV 205
QY 157 LQRRGVP---GGPSGSDPTSRGCVASGNPAILRALDPYKPRRAGFSGESRRRSRAKR 212
Db 206 YVREAVQOEWAEPPDGLHNE-AFGLGDLPLNLLGLVGQD-----LGDTERKRHRAPGS 258
QY 213 FVSIPRYVETLVVADESVMKFGAD-LEHYLLTLTLLATAARLYRHSILNPINIVVVKLL 271
Db 259 Y-----STEVLLVDDSVVRFHGHKEHVONYVITLNNIVDEIYHDESGLGHVNIALVRLIM 313
QY 272 LRDSDGPKV-TGNAALTNRNFCAMQKLNKVSVDKHPYWDTAILFTQDLCGATTCDTL 330
Db 314 VGYRQSLIERGNPSRSLQVCRWAHSQOQDPSHAHHDDHVFLTRQDF-----GPS 367
QY 331 GMADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNMPHD---NVKVCVEEFGKLRAN 387
Db 368 GYAPVTGMCHPLRSCALNHEDGFSFAFIAHETGHVLMGHDGGCGCADETSIGS---- 423
QY 388 HMMSPFTLIQIDRANPWSACSAIITDGLDGHGDCILLDQPSKPI-SLPEDLPGASYTSLQ 446
Db 424 -VMAPLVQAAPHRFHWSKCSKLSRYLPS--YDCLLDDPDPAPWQPPPELPGLGINSME 480
QY 447 QCEALFVGVSKPC---PYMOYCTKLWCTGKAGQMVQOTRHFPHWADGTSCEGKLCCLKGA 503
Db 481 QCRDFGSGYGTCLAFRTFEFCQKQWCS-HPDNPFCKTKGPPDGTGECAPKRWKFGH 539
QY 504 CVERHNLNKHVRDGSWAKWDPYGPCSRTCGGGVQLARRQCNTPTPANGKYCEGVRYK 563

Db 424 -VMAPLVAQAFHRHWSRCSKLELSRILPS--YDCLDDDPFDPAWQPPELPPOINTSMDE 480

Qy 447 QCELAFAVGSKPC---PYMIOYCTKLWCTGAKAGOMWCQTRHFPWADGTCGEGKLCCLKGA 503

Db 481 QCRDFGSGYOTCLAFRTTFPCCKOLMCS-HPDNPYFCKTKKGPLDGTCECAPGKWCFKGH 539

Qy 504 CVERHNLNKHRYDGSWAKWDPYGPCSTCGGVLARQCTNPTPANGKCYCGSVRKVR 563

Db 540 CIWKSPOBTYQDGGGWSWTKFGSCRSKCGGVRKSRSCNNPSPAYGGRCPLGPMFEYQ 599

Qy 564 SCNLEPCPSSASGSKFREEOCEAFNGV-----NHSTNRLTLAVAWVPKYSGVSPDKCK 617

Db 600 VCNSEECPTV--EDFRAQQCAKNSYVHQNAKHS-----WVP-YEPDDDAQKE 647

Qy 618 LICRANGTGYVVLAPKVVDGTICS-PDSTSVCVQKCIKAGCGNLGSKKRRDKCGVCG 676

Db 648 LICQADTGDVVFNNQVYHDTGRTCSYRDPYSVCARGCEVPVGCDEKEYGSMKADDKCGVCG 707

Qy 677 GDNKCKKVTG 687

Db 708 GDNSHCRIVKG 718

RESULT 12

ID Q8TEY8 PRELIMINARY; PRT; 1159 AA.

AC Q8TEY8;

DT 01-JUN-2002 (TReMBLrel. 21, Created)

DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE ADAMTS14.

GN ADAMTS14.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21839041; PubMed=11741898;

RA Collige A., Vandenberghie I., Thilly M., Lambert C.A., Van Beeumen J.,

RA Li S.W., Prockop D.J., Lapiere C.M., Nusgens B.V.;

RT "Cloning and Characterization of ADAMTS-14, a Novel ADAMTS Displaying

RT High Homology with ADAMTS-2 and ADAMTS-3.,"

RL J. Biol. Chem. 277:5756-5766(2002).

DR EMBL; AF366351; AAL79814.1; -

SQ SEQUENCE 1159 AA; 127336 MW; A5B130149BF7FF34 CRC64;

Query Match 24.7% Score 926.5; DB 4; Length 1159;

Best Local Similarity 33.2%; Pred No. 3.1e-68;

Matches 223; Conservative

Qy 45 PEDSGDQGLFIQTAFQEDFYHLHTPDAQFLAPAFSTE-----LGVPLQGLTGSSD 97

Db 31 PGRVGRHSLYFNVFVKELHLRPNRRLVVPSSVWEQDFELRQPLR----- 82

Qy 98 LRRCFYGDVNAEPDSPAAVSLCGLRGATCYRGAEEVISPLPNASAPAAQNSQG-AHL 156

Db 83 -QECVYITGGVTCMPGAVAISNCDGLAGLIRTDSTDFIEPLERQ---QEKASGRTHV 138

Qy 157 LQRGVP----GGPSGDPTSRCGVASGNPAILRALDPYKPRRAGGESRRSRGRAKR 212

Db 139 VYRREAVQOEWAEPDGDHLNE-AFGCLDPLNLLGLVDQ-----LGDTERKKRRHAKPGS 191

Qy 213 FVSPRYVETLVVADESVMKFGHAD-LEHYLLTLLATAARLYRHPSTLININIVVKVLL 271

Db 192 Y-----SIEVLLLVDDSVVRFEKGHEHYQNVYLTLMNTVIDEYHDES LGVHINIALVRLIM 246

Qy 272 LRRDRSGPKV-TGNAALTLRNFCAWQKLNKVSQKHPYWDTAILFTRQDLCGATTCDTL 330

Db 247 VGYRQSLIERGNPSRSLQVCRWAHSQQQDPDFAEHHDHVVLTRQDF-GPSQWQ-- 303

Qy 331 GMADVGTWCDPKRCSVTEDDGLFSATFTTAHELGHVFNMPHD---NVKVCSEVFGKLRA 387

```
Db 304 GYAPVTGMCHPLRSCALNHEHGFSSAFVIAHETGHVLMGHEHDGQNGCADETSLGS----- 359
QY 388 HIMSPTLIQIDRANPWSACSAALITDFLDGSHGDCLLDOPSKPI-SLPEDLPASVTLNQ 446
Db 360 -VNAPLVQAAFRHFWRSCKLSLYLPS--YDCLLDDPFPANPQPELPFGINYSDME 416
QY 447 QCELAFGVSGKPC---PYMOYCTKLWCTGKAKQMVCOQTRHFPWADGTCGSGKLCCLKGA 503
Db 417 QCRFDGSGYQTCIAFRTFEPCKQLWCS-HPDNPYFCKTKGPPDLDTBCAPGKWCFKGH 475
QY 504 CVERHNLNKHRYDGSNAKWDYPGCSRTCGGVQVQLARRQCTNPTPANGKYCEGVRYKXR 563
Db 476 CIWKSPEQYTGQDGWSSWTKFGSCRSRSGVRSRSCNPNPSPAYGGRPCLGPMFEYQ 535
QY 564 SCNLEPCPSSASGKSFREEQCAFNGY-----NHSTNRLTLAVAWVPKYSGVSPRDKCK 617
Db 536 VCNSEECPGTY--EDFRAQCAKRNYSYVHQNAKHS-----WVP-YEPDQDAQCKE 583
QY 618 LICRANGCTGYFVLAIPKVVDTGLCS-PDSTSVCVQCKIKACDGNLGSKKRFRDCKGVG 676
Db 584 LICQSDTGDVFMNQVHDGTRCSYRDPSYVCARGECVPVGCDEVGSMKADCKGVG 643
QY 677 GDNKSCCKKVTG 687
Db 644 GDNSHCRTVKG 654

RESULT 13
O9GL54 PRELIMINARY; PRT; 269 AA.
ID O9GL54
AC O9GL54;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE AggreCanase-2 (Fragment).
GN ADAMTS-11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP Goad D.L., Goad M.E.;
RA "Molecular cloning of lapine aggrecanase-2 (ADAMTS-11) from articular
RT chondrocytes.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF317415; AAC33062.1; -.
DR HSSP; Q9PW35; IBUD.
DR MEROPS; M12.225; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR Pfam; PR01421; Reprolysin; 1.
DR Pfam; PF00090; tsp-1; 1.
DR SMART; SM00209; TSPl; 1.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPl; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1
FT SIGNAL 269
SQ SEQUENCE 269 AA; 29193 MW; 97A1CA80B33452FA CRC64;

Query Match 22.2%; Score 830; DB 6; Length 269;
Best Local Similarity 54.7%; Pred. No. 4.6e-61;
Matches 146; Conservative 37; Mismatches 74; Indels 10; Gaps 3;

QY 329 TLGMADVTGTCMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANH 388
Db 1 TLGMADVTGTCSPERSCAVIEDDGLHAFTVAHEIGHLLGLSHDSDSKFCEENFGSTEDKR 60
QY 389 MMSPTLIQIDRANPWSACSAALITDFLDGSHGDCLLDQPSKPISLPEDLPASVTLNQ 448
Db 61 LMSSILTSIDASKPWSKCTSATITEFLDDBGHGNCLLDVPKQILGPEELPGQTYDATQC 120
```

```
QY 449 ELAFGVGSKPCPYMOYCTKLWCTGKAKQMVCOQTRHFPWADGTCGSGKLCCLKGACVER- 507
Db 121 NLITGPEYTVCPGMDVCARLWCAVVRQGMVCLTKLPAVEGTGCGKRICLOGKCVDKT 180
QY 508 ----HNLNKHRYDGSNAKWDYPGCSRTCGGVQVQLARRQCTNPTPANGKYCEGVRYKXR 563
Db 181 KKTYSTSSH---GNWGSWGPWGQCSRSGCGGVQFAYRHCHNPNAPRNSGRYCTGKRAIYR. 237
QY 564 SCNLEPCPSSASGKSFREEQCAFNGY 590
Db 238 SCSVTPCP--ANGKSFREHQCENKNGY 262

RESULT 14
O9GL37 PRELIMINARY; PRT; 1427 AA.
ID O9GL37
AC O9GL37;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE von Willebrand factor-cleaving protease precursor.
GN ADAMTS13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX PubMed-11557746;
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,
RA Fujikawa K.;
RT "Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a
RT Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.";
RL J. Biol. Chem. 276:41059-41063(2001).
DR EMBL; AY05376; AAL17652.1; -.
DR MEROPS; M12.241; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_Mtpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp-1; 4.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPl; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Signal.
FT SIGNAL 1 33
FT CHAIN 75 1427
FT SIGNAL 1 33
SQ SEQUENCE 1427 AA; 153632 MW; EB1BC3AABCL1A442 CRC64;

Query Match 19.2%; Score 720; DB 4; Length 1427;
Best Local Similarity 31.5%; Pred. No. 6.8e-51;
Matches 179; Conservative 66; Mismatches 240; Indels 84; Gaps 18;

QY 174 CGVAGS-WNP-----AILRALDPY-----KPRRAGFSGSRSSRRSGRAKRF 213
Db 21 CGFLLCWGPVSHFQSCLOALEPQAVSYLSFGAPLKGPPSPGFQQRQR-----RA 75
QY 214 VSIPIRYETLVVADESVMYKFGADLEHYLLTLATAARLYRHPISILPINIVVVKVLLLR 273
Db 76 AGGILHELLELVAVGPDVFAHQEDTERVLTNLNIGABELLRDPSLGAQFRVHLVKMVLIT 135
QY 274 DRDGPVKVTGNAALTLRNFCAWOKLNVSKDHPYWDTAILFTRODL-CGATTCDFLGM 332
Db 136 EPEGAPNTANLTSLLSVCVGSQTINPEDDTDFGHADLVLYITRFDLEDPGNGRVRGV 195
QY 333 ADVGTMCDPKRSQSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 392
Db 196 TQLGACGSPVTSCLITETGFDLGVITAEIGHGSHFGLGHDGAPGS----CGCPSGCHVNAS 251
QY 393 TLIQIDRANP-----WSACSAALITDFLDGSHGDCLLD-----QPSKPISLPEDLPASV 443
```

```
Db 252 -----DGAAPRAGLANSPCSRROLLSLLSAGRARCVMDPPRPQPGSAGHPPDAQPGLIYS 306
QY 444 LSQOCELAFGVGSKPCPY-----MOYCKLMCTGKAKGMVQCOTRHPFPWADGTSCGEGKLC 499
Db 307 ANEQCRVAFGPKAVACTFAREHLDMQALSCHTDPLDQSSCSRLLVPLLDGTGCGVEKWC 366
QY 500 LKAGACVERHNLNK-HRVDSWAKWDYGPCCSRCTCGGGVQLARRQCTNPTPANGGKYCEGV 558
Db 367 SKGRCSLVELTPIAAVHGRWSSGPRSPCSRSCGGGVVTRRRQCNPRPFAFGGRACVGA 426
QY 559 RVKYRSCNLPCCPSSASGSKFRECEQEAENG-----YNHSTNRLTLAVAW---V 604
Db 427 DLAEMCNTQACEKQ--LEFMSQQCARDGQPLRSSPGGASFYH-----WGAAY 474
QY 605 PKYSGVSPDKCKLICRANGTYFYVYLPKPVVDGTLTCLSP-----DST-SVCVQGGKCIKAG 658
Db 475 PHSQGDAAA--LCRHMCRAIGESIMKRGDSFLDGTGRCMPSPGPPREDGTLSLCVSGSCRTFG 531
QY 659 CDGNLGSKKRFDCGVCYGGDNKCKKVTG 687
Db 532 CDGRMDSQQVMDRCQVCGGDNSTCSPRK 560
```

RESULT 15

```
Q95N24 PRELIMINARY; PRT; 192 AA.
AC Q95N24;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Aggrecaanase-1 (Fragment).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Flannery C.R., Little C.B.;
RT "Expression and activity of equine aggrecaanases.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF368321; AAK53425.1; -
DR MEROPS; M12.221; -
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000130; Zn_Mtpeptase.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 192
SQ SEQUENCE 192 AA; 20670 MW; 9013B0E19FCE8C56 CRC64;

Query Match 14.9%; Score 558; DB 6; Length 192;
Best Local Similarity 51.0%; Pred. No. 1.4e-38;
Matches 98; Conservative 26; Mismatches 64; Indels 4; Gaps 3;

QY 332 MADVGTMDCKPSCSVTEDDGLPSAFTTAHELGHVFNMPHDNVKCEVEFGK-LRANHMM 390
Db 1 MADVGTVDCAISCAIVEDDGLQSAFTAHELGHVFNMLHDNSKPCVGLNGPGSTSRHYM 60
QY 391 SPTLIQIDRANPWSACSAALITDLSGHGDCLLDQFSKPISLPEDLPGASYTLTQQCEL 450
Db 61 APVMAHVDPEEPWSPCSARITDLDNGYGHCLLDKPEAPLHPVTFPGKDYDADRQCQL 120
QY 451 AFGVGSKPCPYM-OYCTKLMCTGKAKGMVQCOTRHPFPWADGTSCGEGKLCILKAGACVERHN 509
Db 121 TFGPDSRHCPQLPPPCAAALWCSGHLNGLHAMCQTKHSPWADGTGCPGAQACMGRCCLHMDQ 180
QY 510 LNKHRVD--GSW 519
Db 181 LQEFNIPQAGW 192
```

Search completed: April 29, 2003, 17:17:53
Job time : 45.8928 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: April 29, 2003, 17:12:21 ; Search time 12.9594 Seconds
(without alignments)
3604.758 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_583
Perfect score: 3163
Sequence: 1 MLLGLITLAFAGTAGGFE.....SCNLEPCSSAGSKSFREQ 583

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues
Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PublishedApplications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3155	99.7	950	10 US-09-965-631-4	Sequence 4, Appli
2	3126.5	98.8	823	9 US-10-163-316-2	Sequence 2, Appli
3	1913	50.5	367	10 US-09-965-631-6	Sequence 6, Appli
4	1669	52.8	321	10 US-09-965-631-2	Sequence 2, Appli
5	1562.5	49.4	950	10 US-09-321-987B-4	Sequence 4, Appli
6	1562.5	49.4	968	9 US-10-163-316-7	Sequence 7, Appli
7	1560.5	49.3	967	12 US-10-105-929-2	Sequence 2, Appli
8	1353	42.8	727	9 US-10-097-597-12	Sequence 12, Appli
9	1353	42.8	727	9 US-10-097-580-12	Sequence 12, Appli
10	1351	42.7	727	10 US-09-445-023A-12	Sequence 12, Appli
11	1351	42.7	727	9 US-10-097-597-1	Sequence 1, Appli
12	1351	42.7	727	9 US-10-097-580-1	Sequence 1, Appli
13	1351	42.7	727	10 US-09-445-023A-1	Sequence 1, Appli
14	1337.5	42.3	837	9 US-10-174-590-352	Sequence 352, App
15	1337.5	42.3	837	9 US-10-176-758-352	Sequence 352, App
16	1337.5	42.3	837	9 US-10-175-737-352	Sequence 352, App
17	1337.5	42.3	837	9 US-10-173-706-352	Sequence 352, App
18	1337.5	42.3	837	9 US-10-175-738-352	Sequence 352, App
19	1337.5	42.3	837	9 US-10-175-752-352	Sequence 352, App

ALIGNMENTS

RESULT 1
US-09-965-631-4
; Sequence 4, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-4

Query Match	99.7%	Score 3155;	DB 10;	Length 950;
Best Local Similarity	99.8%;	Pred. No. 9.9e-251;		
Matches 582;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	MLLGLITLAFAGTAGGPEPEREVVPLRLDPDINGRRYYWRGPDSDGQGLIFQITAF	60	
Db	1	MLLGLITLAFAGTAGGPEPEREVVPLRLDPDINGRRYYWRGPDSDGQGLIFQITAF	60	
QY	61	QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRRCFYSGDVNAEPDSFAVSLC	120	
Db	61	QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRRCFYSGDVNAEPDSFAVSLC	120	
QY	121	GGLRGAFYRGAEYVISPLPNASAPAAQNSOGAHLQRRGVPGPGSDPTSRCVAGSW	180	
Db	121	GGLRGAFYRGAEYVISPLPNASAPAAQNSOGAHLQRRGVPGPGSDPTSRCVAGSW	180	
QY	181	NPAILRALDPYKPRAGFGESRRSRRAKRFVSIPIRYVETLVVADFSWKFGADLH	240	
Db	181	NPAILRALDPYKPRAGFGESRRSRRAKRFVSIPIRYVETLVVADFSWKFGADLH	240	
QY	241	YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKLN	300	

20	1337.5	42.3	837	9	US-10-176-482-352	Sequence 352, App
21	1337.5	42.3	837	9	US-10-176-757-352	Sequence 352, App
22	1337.5	42.3	837	9	US-10-176-913-352	Sequence 352, App
23	1337.5	42.3	837	9	US-10-180-552-352	Sequence 352, App
24	1337.5	42.3	837	9	US-10-180-557-352	Sequence 352, App
25	1337.5	42.3	837	9	US-10-173-700-352	Sequence 352, App
26	1337.5	42.3	837	9	US-10-174-572-352	Sequence 352, App
27	1337.5	42.3	837	9	US-10-174-579-352	Sequence 352, App
28	1337.5	42.3	837	9	US-10-174-582-352	Sequence 352, App
29	1337.5	42.3	837	9	US-10-174-588-352	Sequence 352, App
30	1337.5	42.3	837	9	US-10-175-739-352	Sequence 352, App
31	1337.5	42.3	837	9	US-10-175-740-352	Sequence 352, App
32	1337.5	42.3	837	9	US-10-175-743-352	Sequence 352, App
33	1337.5	42.3	837	9	US-10-176-488-352	Sequence 352, App
34	1337.5	42.3	837	9	US-10-176-492-352	Sequence 352, App
35	1337.5	42.3	837	9	US-10-176-747-352	Sequence 352, App
36	1337.5	42.3	837	9	US-10-176-750-352	Sequence 352, App
37	1337.5	42.3	837	9	US-10-176-985-352	Sequence 352, App
38	1337.5	42.3	837	9	US-10-176-987-352	Sequence 352, App
39	1337.5	42.3	837	9	US-10-176-991-352	Sequence 352, App
40	1337.5	42.3	837	9	US-10-176-992-352	Sequence 352, App
41	1337.5	42.3	837	9	US-10-176-993-352	Sequence 352, App
42	1337.5	42.3	837	9	US-10-184-658-352	Sequence 352, App
43	1337.5	42.3	837	9	US-10-173-695-352	Sequence 352, App
44	1337.5	42.3	837	9	US-10-173-697-352	Sequence 352, App
45	1337.5	42.3	837	9	US-10-173-705-352	Sequence 352, App

|||||
Db 241 YLLTLATAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCWQKKLN 300
QY 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
Db 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
QY 361 HELGHVFNMPHDNVKVC EEFVGGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 420
Db 361 HELGHVFNMPHDNVKVC EEFVGGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 420
QY 421 DCLLDQPSKPISLPEDLPAGASYTLSQOCELAFGVSKPCPYMQYCTKLMCTGKAKGQMYC 480
Db 421 DCLLDQPSKPISLPEDLPAGASYTLSQOCELAFGVSKPCPYMQYCTKLMCTGKAKGQMYC 480
QY 481 QTRHFPWADGTSCEGKCLKLGACVERHNLKRVDPGWSAKWDPYPCSRCTCGGGVQLAR 540
Db 481 QTRHFPWADGTSCEGKCLKLGACVERHNLKRVDPGWSAKWDPYPCSRCTCGGGVQLAR 540
QY 541 RQCTNPTPANGKYCEGVRYKRSCLNLEPCPSSASGKSFREQ 583
Db 541 RQCTNPTPANGKYCEGVRYKRSCLNLEPCPSSASGKSFREQ 583

RESULT 2
US-10-163-316-2
; Sequence 2, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MPI01-025PIRNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2

Query Match 98.8%; Score 3126.5; DB 9; Length 823;
Best Local Similarity 95.9%; Pred. No. 1.8e-248;
Matches 581; Conservative 0; Mismatches 2; Indels 23; Gaps 1;
QY 1 MLLLGILTLAPAGTAGGFEPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
Db 1 MLLLGILTLAPAGTAGGFEPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDFAAASLC 120
Db 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDFAAASLC 120
QY 121 GGLRGAFGYRGA EYVISP LNPASAPAAQNSQGAHL LQRRGVPGGSGDPTSRCGVASGW 180
Db 121 GGLRGAFGYRGA EYVISP LNPASAPAAQNSQGAHL LQRRGVPGGSGDPTSRCGVASGW 180
QY 181 NPAILRALDPYKPRRAGFGE SRRSRGRKRFVSI PRYVETLVVADSVK FKGADLEH 240
Db 181 NPAILRALDPYKPRRAGFGE SRRSRGRKRFVSI PRYVETLVVADSVK FKGADLEH 240
QY 241 YLLTLATAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALT LRNFCWQKKLN 300
Db 241 YLLTLATAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALT LRNFCWQKKLN 300
QY 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
Db 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360

QY 361 HELGHVFNMPHDNVKVC EEFVGGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 420
Db 361 HELGHVFNMPHDNVKVC EEFVGGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 420
QY 421 DCLLDQPSKPISLPEDLPAGASYTLSQOCELAFGVSKPCPYMQYCTKLMCTGKAKGQMYC 480
Db 421 DCLLDQPSKPISLPEDLPAGASYTLSQOCELAFGVSKPCPYMQYCTKLMCTGKAKGQMYC 480
QY 481 QTRHFPWADGTSCEGKCLKLGACVERHNLKRVDPGWSAKWDPYPCSRCTCGGGVQLAR 540
Db 481 QTRHFPWADGTSCEGKCLKLGACVERHNLKRVDPGWSAKWDPYPCSRCTCGGGVQLAR 540
QY 518 SWAKWDPYGPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVRYKRSCLNLEPCPSSASGK 577
Db 541 SWAKWDPYGPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVRYKRSCLNLEPCPSSASGK 600
QY 578 SFREQ 583
Db 601 SFREQ 606

RESULT 3
US-09-965-631-6
; Sequence 6, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Fridgele, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 367
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-6

Query Match 60.5%; Score 1913; DB 10; Length 367;
Best Local Similarity 99.7%; Pred. No. 3e-149;
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLLLGILTLAPAGTAGGFEPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
Db 1 MLLLGILTLAPAGTAGGFEPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDFAAASLC 120
Db 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDFAAASLC 120
QY 121 GGLRGAFGYRGA EYVISP LNPASAPAAQNSQGAHL LQRRGVPGGSGDPTSRCGVASGW 180
Db 121 GGLRGAFGYRGA EYVISP LNPASAPAAQNSQGAHL LQRRGVPGGSGDPTSRCGVASGW 180
QY 181 NPAILRALDPYKPRRAGFGE SRRSRGRKRFVSI PRYVETLVVADSVK FKGADLEH 240
Db 181 NPAILRALDPYKPRRAGFGE SRRSRGRKRFVSI PRYVETLVVADSVK FKGADLEH 240
QY 241 YLLTLATAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALT LRNFCWQKKLN 300
Db 241 YLLTLATAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALT LRNFCWQKKLN 300
QY 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
Db 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
QY 361 HELG 364
Db 361 HELG 364

RESULT 4

US-09-965-631-2
; Sequence 2, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Encodir
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 321
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-2

Query Match 52.8%; Score 1669; DB 10; Length 321;
Best Local Similarity 99.4%; Pred. No. 2.8e-129;
Matches 318; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLLGILTLAFAGTAGGFEPEREVVPRLDPDINGRYYWRGPGSDGGLIFQITAF 60
DB 1 MLLGILTLAFAGTAGGSEFEREVVPRLDPDINGRYYWRGPGSDGGLIFQITAF 60
QY 61 QEDFYLHLTPAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDSPAAYSLC 120
DB 61 QEDFYLHLTPAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDSPAAYSLC 120
QY 121 GGLGAFYRGAEEYISPLPNASAPAAORNSQGAHLQRGVPGGSDPTSRGCVASGW 180
DB 121 GGLGAFYRGAEEYISPLPNASAPAAORNSQGAHLQRGVPGGSDPTSRGCVASGW 180
QY 181 NPAILRALDPKPRRAGGESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 240
DB 181 NPAILRALDPKPRRAGGESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 240
QY 241 YLLTLLATAARLYRHPSTILNPINIVVKKVLLLRDRDGGPKVTGNAALTNRNFCAMQKLN 300
DB 241 YLLTLLATAARLYRHPSTILNPINIVVKKVLLLRDRDGGPKVTGNAALTNRNFCAMQKLN 300
QY 301 KVSQKHPEYWDTAILFTROD 320
DB 301 KVSQKHPEYWDTAILFTROE 320

RESULT 5

US-09-321-987B-4
; Sequence 4, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Kimble, Judith E
; APPLICANT: Bieleloch, Robert H
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296.95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/129,023
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Murine

US-09-321-987B-4

Query Match 49.4%; Score 1562.5; DB 10; Length 950;
Best Local Similarity 50.4%; Pred. No. 6.3e-120;
Matches 315; Conservative 79; Mismatches 154; Indels 77; Gaps 16;

QY 1 MLLGILTLAFAGTAGG--FEPERVVVPRLDPDINGRYYWRGP-EDSGGQGLIFQI 57
DB 20 LLLASTITMLLCARGAHGRTEDEELVLP-SLE-----RAPGHDSTTTRL--RL 66
QY 58 TAFQEDFYLHLTPDAQFLAPAFSTEHLGVPLQGLTGGS-----SDLRRCFYSGDVNA 109
DB 67 DAFGQQLHLKLPDPSGFLAPGFTLQIV-----CRSGSEAHQLDPTGDLAHCEYSTVNG 121
QY 110 EPDSFAAVSLCGGLRGAFGYRGAEEYISPLPNAS----APAAORNSQGA----HLQRRG 161
DB 122 DPGSAAALSCEGVGRGAFYLGEEFTIQAPAGVATERLAPAVPEPESSARPOPHILRRR- 180
QY 162 VPGGPGSDPTSRGCVASGWNPAILRALDPYKPRRAGFGESESRRR----- 206
DB 181 ----RRSGGAKCGVMD-----DETLPSTDSPESQNTNRNOMPVRDPTPDAGKP 226
QY 207 ----SGRAKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSTILNP 262
DB 227 SGPGSIRKKRFVSSPRYVETMLVADQSMADFGSGGLKHYLLTLFSVAARFYKHPSTIRSI 286
QY 263 NIIVVKKVLLLRDRDGGPKVTGNAALTNRNFCAMQKLNKVSQKHPEYWDTAILFTRODLC 322
DB 287 SLVVVKILVIYERQKGPETVTSNAALTNRNFCAMQKOHNSPDRDPHEYDTAILFTRODLC 346
QY 323 GATTCTGLMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNDVNVKVEEVEG 382
DB 347 GSHTCTGLMADVTGCDPFRSCSVIEDDGLQAAFTTAHELGHVFNMPHNDVNVKVEEVEG 406
QY 383 KLRANHMSPTLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKPISLPDLPGASY 442
DB 407 VTGDSHLMASMLSDHSPWSPCSAYMTSFLDNGHGECLMDKPNQIKLPDLPGTLY 466
QY 443 TISOQCELAFYGVSKPCP-YMQYCTKLWCTGKAKGMQVQOTRHFHWADGTSQGEKCLK 501
DB 467 DANRQOFTFGEESKHCPCDAASTCTTLWCTGTSGGLLVQCTKHFPWADGTSQGEKWCYS 526
QY 502 GACVERHNLNKH---RVDSWAKWDPYGPCSRTCGGVQOLARRQCTNPTTPANGKYCEGV 558
DB 527 GKCVNKTDM-KHFATPVHSGWGPWGPWDCSRTCGGVQVYTHRECDNPVPKNGKYCEGK 585
QY 559 RVKYSRNCLEPCSSASGKSFREEQ 583
DB 586 RVKYSRNCIEDCPDN-NGKTFREEQ 609

RESULT 6

US-10-163-316-7
; Sequence 7, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552. A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MP101-025P1RNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-163-316-7

Query Match 49.4%; Score 1562.5; DB 9; Length 968;

Query Match 49.3%; Score 1560.5; DB 12; Length 967;
Best Local Similarity 51.6%; Pred. No. 9.4e-120;
Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;

QY 1 MLLGILTLAFAGTAGG--FEPERVVVPIRLDPDINGRRYWRGPDGQGLIFQITAF 60
DB 37 LLLASITLLCAGAHGRTDEELVLP-SLE-----RAPGHDSTTTL--RL 83
QY 58 TAFQDFYHLTPDAQFLAPAFSTEHLGVPLOGLTGGS-----SDLRRFCYSGDVNA 109
DB 84 DAFQGOHLKLPDPSGLAPGFTLQTV-----GRSPGSEAQLHDPDGLAHCFYSGTVNG 138
QY 110 EPDSFAAVSLCGGLRGAFYGAAYVISPLPNAS-----APAAQRNSQGA-----HLLQRRG 161
DB 139 DPGSAALSLCEGRCGAFYLGQEEFFIQAPGVATERLAPAVEEASSARPQFHILRRR- 197
QY 162 VPGSGDPTSRGCVASGWNPAILRALDPYKPRAGFGESRSRR----- 206
DB 198 -----RRSGGAKGCVMD-----DETLPDTSRSPESQNTNQMVPYRDPDQAGKP 243
QY 207 -----SGRAKRFVSIPIRYETLVVVADESVMKFGADLEHYLLTLATAARLYRHPISILNPI 262
DB 244 SGFGSIRKKRFVSPRYETMLVADQSMADFHGSGLKHVLLTLFSVAARFYKHPISIRNSI 303
QY 263 NIVVVKLLLRDRSGPKVTGNAALTLRNFCAWQKLNKVSQKHPEYWDTAILFTRODLG 322
DB 304 SLVVVKILVIYEQKGPEVTSNAALTLRNFCAWQKLNKVSQKHPEYWDTAILFTRODLG 363
QY 323 GATTCDFLGADVTGMDCKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDNVKVCVEVFG 382
DB 364 GSHTCDFLGADVTGMDCKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDNVKVCVEVFG 423
QY 383 KLRANHMSPTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQPSKPSLSPDLPGASY 442
DB 424 VTGDSHLMASLLSDHSQWSPSCAYMTVTSFLDNGHGECLMDKPNQPIKLPSDLPGTLY 483
QY 443 TLOQCELAGVSGKPCP-YMQVCTKLWCTGKAGQWVCOTRHPFWADGTSCEGKLCCK 501
DB 484 DANRQCQFTTEGSEKHCPCDAASTCTTLMWCTGSGGLVLCQTKHFPWADGTSCEGKWCVS 543
QY 502 GACYERHNLNKH---RVDGSMWAKWDPYGPCSRTCGGGVQLARRQCTNPTPANGKCYEGV 558
DB 544 GKCVNKTDK-KHFPATPVHSGKPGWPGWDCSRTCGGGVQVYTMRECDNPVPKNGKCYEGK 602
QY 559 RVKYRSCNLEPCPSSASGKSPREQ 583
DB 603 RVYRSCNIEDCPDN-NGKTFREQ 626

RESULT 7

US-10-105-929-2
; Sequence 2, Application US/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-105-929-2

RESULT 8

US-10-097-597-12
; Sequence 12, Application US/10097597
; Publication No. US20030022352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiji
; APPLICANT: Hakezaki, Michinori
; APPLICANT: Ishioke, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0

```

; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: M
US-10-097-597-1

```

Query Match 42.8%; Score 1353; DB 9; Length 727;
Best Local Similarity 63.3%; Pred. No. 7.2e+103;
Matches 240; Conservative 52; Mismatches 81; Indels

[illegible]

RESULT 9

```

US-10-097-580-12
; Sequence 12, Application US/10097580
; Publication NO. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-097-580-12

```

Query Match 42.8%; Score 1353; DB 9; Length 727;
Best Local Similarity 63.3%; Pred. No. 7.2e-103;
Matches 240; Conservative 52; Mismatches 81; Indels

QY 209 RAKRFVSIPRYVETLWVADESMVKFHGADLEHYLLTLATAARLYRHPHSILNPINIVVK 268

Db	9	KKRFVSSPRYYETMLVADQSMADFHGSGLKHLLTLFVAAAFYKHPISIRNSISLVVVK	68
Qy	269	VLLLRDRDSPKVTGNAALTLRNFCAWQKLNKVSDDKHPEYWDYTAILTFRQDLCCATTC	329
Db	69	ILVIEEQKGPVTSNAALTLRNFCSMQKHNSPDRDPEHYDTAILTFRQDLCSHTCD	129
Qy	329	TLCMADVGIMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKCEEVFGKLRANH	389
Db	129	TLCMADVGIVCDPSRSCSVIEDDGLQAAFTTAHELGHVFNMPHDDAKHCASLNGVSGDS	189
Qy	389	MMSPTLIQIDRANPWSACSAAITDFDLSHGDCLLDOPSKPISLPEDLPGASYTLSDQC	449
Db	189	LMASMLSLDHSQWSPSCSAWYVTSFLDNGHGELMDKPNQPTKLPSDLPTGLYDANRQC	249
Qy	449	ELAFVGSGPCP - YMQYCTKIWCTGKAKQGMVQOTRHFPWADGTSCEGKCLKGACYER	509
Db	249	QFTFGESKHCDDAASTCTTLWCYGTSGGLLVQTKRHPWADGTSCEGKWCYSGKCVNK	309
Qy	508	HNLNKH - --RVDGSNAKWDPYGPCSRCTCGGGVQLARROCTNPTTPANGGKYCEGVKRYRS	568
Db	309	TDM - KFAFTPVHSGWMPGMPGDCSRCTCGGGVQITMRECDNPVKNGGKYCEGKRVYRS	369
Qy	565	CNLEPCSSASGKSFREQ	583
Db	368	CNIEDCPDN - NGKTFREQ	385

RESULT 10

```

US-09-445-023A-12
; Sequence 12, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Ei-ji
; APPLICANT: Hasegaki, Michinori
; APPLICANT: Ishioka, Keliko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 p
; TITLE OF INVENTION: composition and
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/09/445-023A
; CURRENT FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-16042
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-445-023A-12

```

Query Match 42.8%; Score 1353; DB 10; Length 727;
Best Local Similarity 63.3%; Pred. No. 7.2e-103;
Matches 240; Conservative 52; Mismatches 81; Indels 6; Gaps 4

QY	209	RAKREVSIPRYVETLVVADESMVKYFGADLEHYLLTLLATAARLYRHPISILNPINIVVVK	268
Db	9	RKKREVSSPRYVETMLVADQSMADFHGSGGLKYHLLTLLSVAARFYKHPISIRNSISLVVVK	68
QY	269	VLLLRDROSGPKVTGNAALTLRNFCAWQKLNKVKSDKHPYWDOTALTFTRDLCGATTC	328
Db	69	ILVIVEEQKGPVTSNAALTLRNFCSWKQKHNSPSDROPEHYDTALTFTRDLCGSHTC	128
QY	329	TGLMADVTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPDHMKVKVCEEVFGKLRANH	388
Db	129	TGLMADVTCTCDPSKSCSVIEDDGLQAAFTTAHELGHVFNMPHDDAKHCAISLNCVSCDSH	188
QY	389	NMSPTLLIHDRANPWASACSAIITDFLDGSHGDCLLDQSPKISLPEDLPASVTLSSQOC	448

Db 189 LMASMLSSLDHSPWSPCSAYVMTSFLDNGHGECIMDKPQNPILKPLSDLPGLTYDANRQC 248
QY 449 ELAFGVGSKPCP-YMOYCTKLMCTGKAKQOMVQTRHFPWADGTSCEGKLCGLKAGCYER 507
Db 249 QFTFGESKHCDDAATCTTLWCTGTSGLLVQCQKHFPWADGTSCEGKWCVCVKCNK 308
QY 508 HNLNKH---RVDSWAKWDPYPCSTCGGGVQLARRQCTNPTPANGKYCEGVRYKRS 564
Db 309 TDM-KHFATPVHSGWGPWPGWDCSRTCGGGVQVYTMRECDNPVPKNGKYGCEGKRVYRS 367
QY 565 CNLEPCPSSASGSKSFREQ 583
Db 368 CNLEDCPDN-NGKTFREQ 385

RESULT 11
US-10-097-597-1
; Sequence 1, Application US/10097597
; Publication No. US20030022352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakezaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-597-1

Query Match 42.7%; Score 1351; DB 9; Length 727;
Best Local Similarity 63.6%; Pred. No. 1e-102;
Matches 241; Conservative 50; Mismatches 82; Indels 6; Gaps 4;
QY 209 RAKRFVSIPRYVETLVVADESVMKFGADLEHYLLTLLAARLYRHPSTILNPINIVVVK 268
Db 9 RKKRFVSSPRVYETMLVADQSMAEFHGSLKHYLLTLFSAARLYKHPSIRNSVSLVVK 68
QY 269 VLLLRDRDSGPKVTGNAALTLRNFCWQKLNKVSQKHPEYWDTAIFTRQDLGATTC 328
Db 69 ILVHDEQKGPVTSNAALTLRNFCWQKHNPSPDRDAEHYDTAIFTRQDLGSGTCD 128
QY 329 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANH 388
Db 129 TLGMADVGTVCDDPKRSCSVIEDDGLQAFTTAHELGHVFNPHDDAKQCASLNGVND 188
QY 389 MMSPTLIQIDRANPWSACSAAITDLDGSHGDCLLDQSPKPSLPEDLPGASVYLSQOC 448
Db 189 MMASMLSLNLDHSPWSPCSAYMITSFLDNGHGECIMDKPQNPILKPLSDLPGLTYDANRQC 248
QY 449 ELAFGVGSKPCP-YMOYCTKLMCTGKAKQOMVQTRHFPWADGTSCEGKLCGLKAGCYER 507
Db 249 QFTFGESKHCDDAATCTTLWCTGTSGLLVQCQKHFPWADGTSCEGKWCVCVKCNK 308
QY 508 HNLNKH---RVDSWAKWDPYPCSTCGGGVQLARRQCTNPTPANGKYCEGVRYKRS 564
Db 309 TDM-KHFATPVHSGWGPWPGWDCSRTCGGGVQVYTMRECDNPVPKNGKYGCEGKRVYRS 367

QY 565 CNLEPCPSSASGSKSFREQ 583
Db 368 CNLEDCPDN-NGKTFREQ 385
RESULT 12
US-10-097-580-1
; Sequence 1, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakezaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-580-1

Query Match 42.7%; Score 1351; DB 9; Length 727;
Best Local Similarity 63.6%; Pred. No. 1e-102;
Matches 241; Conservative 50; Mismatches 82; Indels 6; Gaps 4;
QY 209 RAKRFVSIPRYVETLVVADESVMKFGADLEHYLLTLLAARLYRHPSTILNPINIVVVK 268
Db 9 RKKRFVSSPRVYETMLVADQSMAEFHGSLKHYLLTLFSAARLYKHPSIRNSVSLVVK 68
QY 269 VLLLRDRDSGPKVTGNAALTLRNFCWQKLNKVSQKHPEYWDTAIFTRQDLGATTC 328
Db 69 ILVHDEQKGPVTSNAALTLRNFCWQKHNPSPDRDAEHYDTAIFTRQDLGSGTCD 128
QY 329 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANH 388
Db 129 TLGMADVGTVCDDPKRSCSVIEDDGLQAFTTAHELGHVFNPHDDAKQCASLNGVND 188
QY 389 MMSPTLIQIDRANPWSACSAAITDLDGSHGDCLLDQSPKPSLPEDLPGASVYLSQOC 448
Db 189 MMASMLSLNLDHSPWSPCSAYMITSFLDNGHGECIMDKPQNPILKPLSDLPGLTYDANRQC 248
QY 449 ELAFGVGSKPCP-YMOYCTKLMCTGKAKQOMVQTRHFPWADGTSCEGKLCGLKAGCYER 507
Db 249 QFTFGESKHCDDAATCTTLWCTGTSGLLVQCQKHFPWADGTSCEGKWCVCVKCNK 308
QY 508 HNLNKH---RVDSWAKWDPYPCSTCGGGVQLARRQCTNPTPANGKYCEGVRYKRS 564
Db 309 TDM-KHFATPVHSGWGPWPGWDCSRTCGGGVQVYTMRECDNPVPKNGKYGCEGKRVYRS 367
QY 565 CNLEPCPSSASGSKSFREQ 583
Db 368 CNLEDCPDN-NGKTFREQ 385

RESULT 13
US-09-445-023A-1
; Sequence 1, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka


```
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-352

Query Match      42.3%; Score 1337.5; DB 9; Length 837;
Best Local Similarity 45.1%; Pred. NO. 1.6e-101;
Matches 265; Conservative 88; Mismatches 192; Indels 43; Gaps 12;

Qy 1 MLLGLITLAFAGTGGFEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFOITAF 60
Db 37 LLLLLASLLPSARLASPLREEIYFPEKLVSVL-----PCSGAPARLLCRQAF 88
Qy 61 QEDFYHLTPDAQFLAPAFSTEHLGVLQGLTGGSSDLRRCFYSGDVNAEPDSFAAVSLC 120
Db 89 GETLLLEQDSGVQVVEGLVQLGQAPE-LLGGAEP--GTYLGTINGDPESVASLHWD 145
Qy 121 GG-LRGAFGRGAYIVISPLPNASAPAAQRNSOGAHLQRRGVPGGPDPTSRGCVASG 179
Db 146 GGALLGVLYQYRGAEHLQLPLEGGTPNSA--GGPGAHLRRK----SPASQGGPMCNV--- 196
Qy 180 WNPAILRALDPYKPRRAGFESRRRSRGRKRFVPIRYVETLVVADESVMVKFHGADLE 239
Db 197 -----KAPLGSPSPRR--RAKRFASLSRFVETLVVADDKMAAFHGAGLK 239
Qy 240 HYLTLTLAARLYRHPHSILNPINIVVVKVLLLRDRDSGPKVTGNAALTILRNFCAMQKKL 299
Db 240 RYLLTVMAAAKAFKHPHSIRNPVSLVTVRLVILSGEGEPQVGPSSAAQTLRSFCAMQRL 299
Qy 300 NKVSDKHPEYWDTAILFTRODLGATTCDTLGMADVGTMCDPKRSQSVIEDDGLPSAFTT 359
Db 300 NTPEDSGPDHFDTAILFTRODLGCVSTCDTLGMADVGTVCDPARSCAIVEDDGLQSAFTA 359
Qy 360 AHELGHVFNPHDNVVKYCEEVFGKL-RANHMSPTLIQIDRANPWSACSAAIITDFLDG 418
Db 360 AHELGHVFNMLHDSKPCISLNGPLSTRHVMAVPVMAHVDFEPWSPCSARFITDFLDNG 419
Qy 419 HGDCLLDQPSKPSLPELPGASYTLSSQCELAFGVSKPCPYM-QYCTKLWCTGKAKGQ 477
Db 420 YGCLLDKPEAPLHLPVTFPGKDYADROQCLTFGPDSDRHCQPLPPPPCAALWCSGHLNGH 479
Qy 478 MVCQTRHFPWADGTSCEGKCLKLGACVERHNLNKHVVD--GSWAKWDYPYPCSRCTCGG 535
Db 480 AMCQTKHSPWADGTPCGPAQCMGGRCGLHMDQLQDFNIPAGGWPMPGWDGCSRTCGGG 539
Qy 536 VLARRQCTNPTPANGKGYCEGVKRYSCNLEPCPSSASGKSFREQ 583
Db 540 VQFSRDRCTRPVPRNGKGYCEGRTRFRSCNTEDCP-TGSALTFRREQ 586
```

Search completed: April 29, 2003, 17:21:53
Job time : 16.9594 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 13 4848 Seconds
(without alignments)
4156.253 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_583

Perfect score: 3163

Sequence: 1 MLLGLILTLAFAGRTAGGFE.....SCNLEPCPSASGKSFREQ 583

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1562.5	49.4	951	2 T00017	gene ADAMTS-1 prot
2	1336.5	42.3	837	2 T00355	hypothetical prote
3	941.5	29.8	2165	2 T21371	hypothetical prote
4	775	24.5	1205	2 T18517	procollagen N-endo
5	684	21.6	550	2 T47158	hypothetical prote
6	433	13.7	1444	2 T18856	angiogenesis inhib
7	393	12.4	860	2 T16892	hypothetical prote
8	336	10.6	957	2 T15976	hypothetical prote
9	329	10.4	903	2 S60257	meltrin alpha - mo
10	309	9.8	617	2 S48160	metalloproteinase
11	307	9.7	411	1 HYSNPA	fibrolase (EC 3.4.
12	306	9.7	571	2 S24789	jararagin C precu
13	304	9.6	407	2 S66260	metalloproteinase
14	302.5	9.6	609	2 S55270	catalocollastatin p
15	301	9.5	826	2 A60385	monocyte surface a
16	297	9.4	480	1 A30065	trigamin precursor
17	297	9.4	549	2 S48169	metalloproteinase
18	292	9.2	478	2 JC4880	fibrinolytic metal
19	290.5	9.2	481	2 JC4342	fibrinolytic prote
20	289	9.1	414	2 S41609	atrolysin C (EC 3.
21	284	9.0	414	1 HYSRAC	atrolysin C (EC 3.
22	281	8.9	789	2 S28259	androgen-regulated
23	280	8.9	481	2 S43125	triamucin precursor
24	273.5	8.6	478	2 A43296	atrolysin E (EC 3.
25	268	8.5	616	2 A55796	ecarin precursor -
26	265	8.4	414	2 S41608	atrolysin B (EC 3.
27	255	8.1	610	2 JC7530	vascular apoptosis
28	255	8.1	814	2 G02390	disintegrin-like m
29	243.5	7.7	776	2 S28258	androgen-regulated

30	243	7.7	478	2 JQ1301	hemorrhagic protei
31	231.5	7.3	508	2 T22836	hypothetical prote
32	229.5	7.3	1170	1 TSHUP1	thrombospondin 1 p
33	224.5	7.1	1170	2 A40558	thrombospondin 1 p
34	218.5	6.9	419	2 S41607	atrolysin A (EC 3.
35	217.5	6.9	419	2 A59414	metalloproteinase
36	215.5	6.8	825	2 S55060	feritin alpha-II
37	215	6.8	670	2 T65967	disintegrin-like m
38	213.5	6.7	1042	2 T26644	hypothetical prote
39	210	6.6	416	2 A37877	hemorrhagic protei
40	210	6.6	524	2 S38539	disintegrin-like m
41	202.5	6.4	823	2 S18968	cyritestin precurs
42	200.5	6.3	756	2 S47656	TMDC II protein -
43	200	6.3	952	2 T18900	disintegrin and me
44	197	6.2	1074	2 JC5928	semaphorin F precu
45	196.5	6.2	905	2 S55059	feritin alpha-I -

ALIGNMENTS

RESULT 1

T00017

gene ADAMTS-1 protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00017

R;Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.

Genomics 46, 466-471, 1997

A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1

A:Reference number: Z14055; MUID:98110583; PMID:9441751

A:Accession: T00017

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-951 <XUN>

A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057

A:Experimental source: strain 129SVJ

C:Genetics:

A:Gene: ADAMTS-1

A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2

C:Superfamily: thrombospondin type 1 repeat homology

F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 49.4%; Score 1562.5; DB 2; Length 951;
Best Local Similarity 50.4%; Pred. No. 7.4e-110;
Matches 315; Conservative 79; Mismatches 154; Indels 77; Gaps 16;

QY	1	MLLLGILTLAFAGRTAGG--FEPEREVVVPIRLDPDINGRRYYWRGP-EDSGDQGLIFOI	57
Db	20	LLLLASITMLLCARGAHRPTDEDELVP-SLE-----RAPGHDSTTRL--RL	66
QY	58	TAFQEDFYHLPTDPAQFLAPAFSTEHLGVPLOGLTGGG-----SDLRRCFYSGDVNA	109
Db	67	DAFGQQLHLKLOPDSGLFAPGFTLTQV-----GRSPSEAQHLDPDGLAHCFYSGTVNG	121
QY	110	EPDSEFAVSLCGLRGAFYRGAEVVISPLPNAS----APAAQRNSQGA----HLLQRRG	161
Db	122	DGSAALSLCGVRGAFYLGQEEFFIQAPGVATERLAPAVPEESSARPOFHILRRR-	180
QY	162	VFGPGSGDPTSRGVSAGWNPAILRALDPYKPRRAGFGESRRRR-----	206
Db	181	----RRSGGAKCGVMD-----DETLTSDSRPESQNTNRQWVRDTPDQAGKP	226
QY	207	----SGRAKRFVSIPIRYVTLVVADESVMKFGADLEHYLLPLATAARLYRHPISILNPI	262
Db	227	SGPGSIRKKRFVSSPRYVETMLVADQSMADFHGSLKHYLLTFLFVAAFRYKHPISIRSI	286
QY	263	NIWVKVLLLRDSDGPKVTGNAALTNRNFCWQKLNKVSQDKHPHYMDTALLFTRODLC	322
Db	287	SLVWVKILVIYEQGPEVTSNAALTNRNFCWQKOHNSPSDRDPEDHYDTALLFTRODLC	346
QY	323	GATTCDTLGMADVGTMCDFKRCSCVIEDGLPSAFTTAHELGHVFNMPHDNVKVEEVEFG	382

Db 347 GSHTCDTLGMADVGTCDPSRSCVIEDDGLQAFTTAHELGHVFNMPHDDAKHCASLNG 406
QY 383 KLRANHMSTLTIQIDRANPWSACSAAITDFLDSGHGCLLDOPSKPTISLPEDLPAGSY 442
Db 407 VTGDSHLMASLSLSDHQSQWSPSCSAIYMTFSDNHHGCLMDKPNPKLPSDLPCTLY 466
QY 443 TLSQOCFLAFGVGSKPCP-YMOYCTKLWCTGKAGQWQVQTRHPFWADTSCGEGKLCUK 501
Db 467 DANRQCOFTGESKHCPCDAASCTTLMCTGTSGGLLVQCTKHPFWADTSCGEGKWCYS 526
QY 502 GACVERHNLNKH---RVDSGNKWDYGPSCSTCGGQVQLARRQCTNPTPANGKKYCEGY 558
Db 527 GKCVNKTDM-KHEATPVHGSWGPWGDGRTCGGQVQVTMRCDNPVPKNGKYCEGK 585
QY 559 RVYRSCNLEPCSSASGKSFREEQ 583
Db 586 RVYRSCNLEPCDN-NGKTFREEQ 609

RESULT 2
T00355
hypothetical protein KIAA0688 - human
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00355
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: Z14142; MUID:98403880; PMID:9734811
A:Accession: T00355
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-837 <ISH>
A:Cross-references: EMBL:AB014588; NID:G3327189; PIDN:BAA31663.1; PID:G3327190
A:Experimental source: Brain
C:Genetics:
A:Gene: KIAA0688
C:Superfamily: thrombospondin type 1 repeat homology
F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 42.3%; Score 1336.5; DB 2; Length 837;
Best Local Similarity 45.1%; Pred. No. 7.3e-93;
Matches 265; Conservative 88; Mismatches 192; Indels 43; Gaps 12;

QY 1 MLLGILTLAGTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDGGLIFQITAF 60
Db 37 LLLLLLASLLPSARLASPLREEIIVPEKLNGSVL-----PGSGTPTARLLCRLOAF 88

QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVLPGLTGGSSDLRCFCYSDYNAEPDFAAVSLC 120
Db 89 GETLLEQDSGVQVEGLTVQYLGQAPE-LLGGAEP--GTYLGTINGDPESVASLHWD 145

QY 121 GG-LRGAFGYRGAEVVISPLNAPAAQRNSQGAHLQRRGVPGPGSGDPTSRGCVASG 179
Db 146 GGALLGVLYRGAELHLQPLEGGTPNSA--GGPGAHILRRK-----SPASGGQPMCNV--- 196

QY 180 WNPAILRALDYPKPRAGGSGRRSRRSGRAKRFVSPRYVETLVVADESVMKFGHADLE 239
Db 197 -----KAPLGSPSPRPR--RAKRFASLSRFVETLVVAADDKMAAFHAGLK 239

QY 240 HYLLTLLATAARLHYPHSILNIPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWOKKL 299
Db 240 RYLLTVMAAAKAFKHPISRPVSLVLRVLGSGEGEPQVGSAAQTLSFCAWORGL 299

QY 300 NKVSDKHPYWDTAIFTRQDLCAATCTDLGMADVGTMCDDPKRSCSVIEDDGLPSAFTT 359
Db 300 NTPEDSDPHEDTAIFTRQDLGVSTCDTLGMADVGTVCDDPARSCAIVEDDGLQSAFTA 359

QY 360 AHELGHVFNMPHNDNVKVEEFGKL-RANHMSPTLIQRANPWSACSAAIITDFLDSG 418
Db 360 AHELGHVFNMLHDNSKPCISLNGPLSTSRHVMAPYMAHVDPPEEPWSPCSARFITDLDNG 419

QY 419 HGDCLLDQSPKPSLPELPGASYTLSSQCELAFCVGSKPCPYM-QYCTKLWCTGKARGQ 477

Db 420 YGHCLLDPEAPLHLPTVFPKDYADRCQCLTFGPDSSRHCQPLPPCAALWCSGHLNGH 479
QY 478 MVCOTRHPWPADGTSCGEGKLCGLKACACVERHNLNKHRYD--GSWAKWDYGPSCSRTCGGG 535
Db 480 AMCOTKHSFWADGTPCGPAQACMGRCRLHMDQLQDFNIPQAGGWPWGPWGDSCRTCGGG 539
QY 536 VOLARROCTNTPANGKYCGVRYKYSNLEPCPSSASGKSFREEQ 583
Db 540 VQFSSRDCTRPVPRNGKYCEGRTFRSCNTEDCP-TGSALTFREEQ 586

RESULT 3
T21371
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21371; T24896
R:Gajadsty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone F25H8
R:Gajadsty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
C:Genetics:
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1;

Query Match 29.8%; Score 941.5; DB 2; Length 2165;
Best Local Similarity 38.7%; Pred. No. 1.6e-62;
Matches 203; Conservative 75; Mismatches 181; Indels 65; Gaps 14;

QY 101 CFYSGDVNAEPDFAAVSLC---GGLRGAFGYRGAEVVISPLNPSA---PAAQRNSOGA 154
Db 169 CIYRAHVKG-VHQHSTVNLCDSEGLYGLMALPSGIHTVETPIISGNGTEHDGASHRQ--- 225

QY 155 HLQRRGVPGSGDPTSRGCV-----ASGNPAILRALDYPKPRRAGFSGRSR 204
Db 226 HLVRKFDPMHFKSFDHLNSTSVNETTETTVATWQDQWEDVIER-----KARSR 272

QY 205 RRSGRARFVSPRYVETLVVADESVMKFGHADLEHYLLTLLATAARLHYPHSILNIPINI 264
Db 273 RAAN-----SWDHYVEVLVWADTKMYEYHGRSLEDYVLTFTVASIYRHOSLRASIN 326

QY 265 VVYKVVLLLRDRDSGPKVTGNAALTLRNFCAWOKKLNVKSDKHPEYWDTAIFTRQDLCA 324
Db 327 VVYKLVLTENAGPRITQNAOQTLQDFCRWQYYNDPDDSSVQHHIDVAILLTRKDCIRS 386

QY 325 T-TCDTIGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAAHELGHVFNMPHNDNVKVEE--- 380
Db 387 QGKCDTLGLAELTMDCMQKSCAIIEDNGLSAAFTIAHELGHVFSIPHDERKCSYMPV 446

QY 381 -----FGKLRAN---HMSPTLIQIDRANPWSACSAAIITDFLDSGHG--DCLLD 425
Db 447 NKVCKFGSTKFDKTFQNNFNHINAPTLEYNTHPWSWSPCSAGMLERFLENRRGQTCLFD 506

QY 426 QPSKPSLPE---DLPGASYTLSSQCELAFCVGSKPCPYM-QYCTKLWCTGKARGQVQ 481
Db 507 QPERRYIEDVFRDEPGKIDAHQCKFVFGPASELCFPMPTCRLWCATFVGSQMGCR 566

```

Db      524  KGPLDGTWCAPGKHCFCGHCIWLTPDIILKR-----DGNWGANSPFGSCSRTCTGV 575
               |||:|||||:|||||:::|||||:|||||||:|
Qy      537  QLARRCQTNPTPANGKYCEGRVRYRSCNLPCFPSSASGKSFRREQ 583
               |||:|||||:|||||:::|||||:|||||||:|
Db      576  KFRTRQCNDPNHPANGGRTCSGLAYDLQCLNSODCPDALA--DFREEQ 620
               |||:|||||:|||||:::|||||:|||||||:|

RESULT 5
T47158
hypothetical protein DKFzp762cll110.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47158
R;Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24379
A:Accession: T47158
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-550 <AAA>
A:Cross-references: EMBL:AL162080
A:Experimental source: adult melanoma (MeHo cell line); clone DKFzp762cll110
C:Genetics:
A>Note: DKFzp762cll110.1

Query Match          21.6%; Score 684; DB 2; Length 550;
Best Local Similarity 59.2%; Pred. No. 8.le-44;
Matches 119; Conservative 25; Mismatches 51; Indels 6; Gaps

Qy      387  NHMSPFLIIDRANPWASACSAIIITFDLDHGHGDCLLDPQSKPSILPEDLPGASYTLQS 446
               :|||:|||||:|||||:::|||||:|||||||:|
Db      10  SHMWASMLSLNDHSQPWSQAYMITSFLDNGHGECIMDKPONQLPGDLPGTSYDANR 69
               :|||:|||||:|||||:::|||||:|||||||:|
Qy      447  QCELAFGCGSKPCP-YMQYCTKLWCCTGAKGOMVCOTRHFPWADGTSCGEGLCKLGACV 505
               |||:|||||:|||||:::|||||:|||||||:|
Db      70  QCQFTTGEDSKHCPDDAASTGSTLWCCTGSGVLVCOTKHFPWADGTSCGEKWINGKCV 129
               |||:|||||:|||||:::|||||:|||||||:|
Qy      506  ERHNLANKH---RYVDGWAKWDYPGPCSRTCGGGVQLARROCTNPTPANGKYCEGRVKY 562
               |||:|||||:|||||:::|||||:|||||||:|
Db      130  NKTD-RKHEDTFPHGSWMGMWPWGDCSRCTCGGVGYTMRECDNPVPKNKKYCEGKRVRY 188
               |||:|||||:|||||:::|||||:|||||||:|
Qy      563  RSCNLEPCPSSASGKSFRREQ 583
               |||||:|||||:::|||||:|
Db      189  RSCNLEDCPDN-NKGTFREQ 208
               |||||:|||||:::|||||:|

RESULT 6
T18856
angiogenesis inhibitor homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18856; T24653
R;McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19031
A:Accession: T18856
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WIL>
A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN000028; CESP:C02B4.1
A:Experimental source: clone C02B4
R;McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19917
A:Accession: T24653
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1444 <W12>
A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN000028; CESP:C02B4.1
A:Experimental source: clone t07C5
C:Genetics:
A:Gene: CESP:C02B4.1

```


Db 276 IRYEVLTRPSALAGLHNGHNAOMYLDRCYQRNL-AVRD-----WDHAIMLTGYDIH 329
QY 322 CGATTCTDLGMADVTGMDPKRSGSVIEDDGLPSAFTTAHELGHVFNPHDNVAVKCEVF 381
Db 330 RGAGRSISGTARLDGMDPNTCULABGLDFTSAFTHLGHHRFTLKD----- 381
QY 382 KKLRAHNMSPYL-----IQIDRANPWSACS-----AALIT-----DFLDSGHG 420
Db 382 ---KSDTLTGTFCGSPKWCGLGRCVPWTGNEIQPTVQHVAPVTVTLPSRIDGSWSGW 438
QY 421 DCLLDQ-----PSKIPSLPEDLPGAS-----YTLSSQCELA----- 451
Db 439 ATICSQCTCNGILSGVGLAIARTCSAPYPANGGSDCVGSTRAVLCSROCGRASKSVDE 498
QY 452 -----FGVSGPCPYMVOYCTKLWCTGKAKGOMVQTRHFPW-----A 488
Db 499 YISDKMBQRLKNDRELTHGSGOLNRPFPQACKVFCDVQ---OHYGSQRYRFFGNL 555
QY 489 DGTSGGEGKCLKGACVERHNLKHRVD-----GSAKWDPYPG 527
Db 556 DGTSGVDYCLDGECLALNCCNALISRDQSCPTDTCPTDQSSSVYRGOWGTWSLWTS 615
QY 528 CSRTCGGVLARQCTNPTPANGKVCCEGVYKYRSCNLEPCPS 572
Db 616 CTATCGGYRNRNACS-----ITG--QCEGNEDETEVCSSCPS 654

RESULT 9
S60257
metlin alpha - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
C:Accession: S60257
R:Yagami-Hiromasa, T.; Sato, T.; Kurisaki, T.; Kamiyo, K.; Nabeshima, Y.I.; Fujisawa-Seh
Nature 377, 652-656, 1995
A:Title: A metalloprotease-disintegrin participating in myoblast fusion.
A:Reference number: S60257; MUID:96026308; PMID:7566181
A:Accession: S60257
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-903 <VAG>
A:Cross-references: EMBL:D50411; NID:g1054586; PIDN:BA08912.1; PID:g1054587
C:Superfamily: mouse metlin alpha; disintegrin homology
F:421-503/Domain: disintegrin homology <DIS>
F:349/Active site: Glu #status predicted

Query Match 10.4%; Score 329; DB 2; Length 903;
Best Local Similarity 21.5%; Pred. No. 9.4e-17;
Matches 159; Conservative 74; Mismatches 245; Indels 262; Gaps 27;

QY 2 LLLGILFLAFAGRTA-----GGFEPEREVVPVIRLDPDINGRRYWRGPDSDGQG 52
Db 15 LLLALAGALLAPRAAGMSLWDQRGAVEVARASL--LSKDPGPGQSI---PAKDHPDV 68

QY 53 LIQITAFQEDFVHLTPDAQFLAPAFSTHGLVPLQ-----CLTGGSDLRRCFYSGDV 107
Db 69 LTVOLQLESRLILSLERNEGLTANGFTETHY---LDQGTVDVSLTRNHTD--HCYVHGHW 123

QY 108 NAEFDSFAAYSLCGLRGAFYRGAEYVISPLNPNAS-----APA-AQRNSQGAHLQRG 161
Db 124 QGDAASVVLSTCSDLRGLIMEFNKYSLPEMKNNTDSYKLYPAESKNTIQL----- 176

QY 162 VPGPSPDPTSRGCVAGSNPAILRALDPYKPRRAGFSGRSRRSRGKRVFSIPRYVE 221
Db 177 -----CGSHNKSNTLMEVDSP-----GTSQMRARRHK---RETLMKTKYVE 215

QY 222 TLVVADESMVKFPGADLEHYLLTLLATAARLYRHPISILNPINIVYKVLILLDRDSGPV 281
Db 216 LVIVADNREFORQKLEKVKQRLIAHVDKFFPLN-IRVLGVGVWNDIDK-CST 273

QY 282 TGNALTLRNFCAWOKKLVSKDHPYWDYTDAILFTRQDLGATTCDTLGMADVGTMDCP 341
Db 274 SQDPFTRLHEFLDWRKIKLLPRKSH---DNAQLISGVYFQGT---TIGMAPIMSMCTA 325

QY 342 KRSCSVI---EDDGLPSAFTTAHELGHVFNPHDNV-KVCEEVFGKLRANHMSPYLIQI 397
Db 326 EQSGVYVMDHSDSPGLGAAVTTLAHELGHNFNMHDTLERGCSCRMAAEKGCCIMNPS---- 381
QY 398 DRANP---WSACSAALITDFLDSGHGDCLLDOP----- 427
Db 382 -TGFPFPMVFFSSCRKDLASLEKGMGMCLFNLPEVKQAFGGRKCGNGYVERGEEDCGE 440
QY 428 -----CTKLWCTGKAK-----SKPISLPEDLPGAS--- 441
Db 441 PEBCTNRCCNATTCTLKPDVAHAGOCCEDCQLKPPGTACTACRGSSNSCDLPFECTGTAPHC 500
QY 442 -----YTLSSQCELAFCVGSKPCPYMAY----- 464
Db 501 PANVYLHDGHPGQVDGYCYNGICQTHEQCQVTLWMPGAKPAPGICFERVNSAGDPYGC 560
QY 465 -----CTKLWCTGKAK-----GOMVQTRHFPW 487
Db 561 GRDSKSAFAKCELRDAKCKIOCGGASRPVIGTNNVSIETNIPQEGGRILCRGTHVYL 620
QY 488 AD-----GTSCGEGKCLKGACVERHNLKHRVDGWSWAKWDPYGPCSRTC--GGGV 536
Db 621 GDMPPDGLVLGATKCAEGKICLNRRCONISVFGVHK-----CAMQCHGRGV 667

QY 537 QLARRQCTNPTPANGKGYCE 556
Db 668 CNRNKNC-----HCE 677

RESULT 10
S48160
metalloproteinase (EC 3.4.24.-) H-I precursor - carpet viper
N:Contains: disintegrin
C:Species: Echis pyramidum leakeyi
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C:Accession: S48160
R:Palmer, M.J.I.; Moura-Da-Silva, A.M.; Theakston, R.D.G.; Crampton, J.M.
Eur. J. Biochem. 224, 483-486, 1994
A:Title: Cloning of metalloproteinase genes in the carpet viper (Echis pyramidum leakeyi)
A:Reference number: S48160; MUID:95010025; PMID:7925363
A:Accession: S48160
A:Molecule type: mRNA
A:Residues: 1-617 <PAT>
A:Cross-references: GB:X78970; NID:g763092; PIDN:CAA55565.1; PID:g763093
C:Superfamily: mouse metlin alpha; disintegrin homology
C:Keywords: hydrolase; metalloproteinase; venom
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-192/Domain: propeptide #status predicted <PRO>
F:193-617/Product: metalloproteinase H-I #status predicted <WAT>
F:403-485/Domain: disintegrin homology <DIS>
F:337/Active site: Glu #status predicted

Query Match 9.8%; Score 309; DB 2; Length 617;
Best Local Similarity 25.3%; Pred. No. 1.9e-15;
Matches 119; Conservative 60; Mismatches 206; Indels 86; Gaps 17;

QY 45 PEDSGDGLIQTAFQEDFVHLTPDAQFLAPAFSTHGLVPLQGLTGGSDLRRCFY 104
Db 50 PEQKYEDTMRVEFKVNGEPVVLHLEKNKGLFSEDYSETHYSPDGREITNPPVEDHCYH 109

QY 105 GDVNAEPDSFAAYSLCGLRGAFYRGAEYVISPL--PNASAPAAQRNSQGAHLQRG 162
Db 110 GRIQNDADSTASMSACNLGKGYFMLRGETYLLIEPLKIPDSEAHAVKYENVE----- 161
QY 163 PGGPSGDPSTRCGVA--SCWNPAILRALDPYKPRRAGFSGRSRRSRGKRVFSIPRYVE 221
Db 162 ---KEDEAPKMGVQTWNES-----DELKASOLVATSEQR-----SVYKYIE 203

QY 222 TLVVADESMVKFPGADLEHYLLTLLATAARLYRHPISILNPINIVY-----VKVLLLR 273
Db 204 FVVVADYIMRYKNNND-----STAVRRRIYEIVNLNMVYIVFNHVALTHIEIWSTR 256

Qy 274 DRDGPVTVGNAALTLRNFCAWOKLNVSKHPEYWDTAIFLTRODLGCATTCDTPLGWA 333
 :
 Db 257 DQIT---VQSADVTLDLFEGDWRAKNLLTRKKH-----DNAQLFTGINLG----QTGLIA 305
 Qy 334 DVGTACDKPKRSCTVED---DGLPSAFTTAHELGHVFENPHDNVKV-CREEFGKLRANHM 389
 :
 Db 306 RMGMCMSPNSGVGIQDYCKNLLVAITWAHELGNLGHNDNGNCNCPTSCIMSAVAG 365
 Qy 390 MSPTLTIQIDRANPWSACSAAIITFDLDSHGDCCLLDQPSPKISLPEDLPASVYT-LSQQOC 448
 :
 Db 366 PEPVE-----SFSNCRSDRYSRFNSQSKCIDNKPLKTIVSPVCNGTVEVGEEC 418
 Qy 449 ELAFVGSGKPMPVMQCYTKLWGCGAKAGQMVCOTRFHPWADGTFSCGGKLC 499
 :
 Db 419 D-----CGSRTYCRRNPCCNA-----TTCKL-----TPGSQCADGECC 450

RESULT 11
 HVSNTA
 fibrolase (EC 3.4.24.-) precursor - southern copperhead
 C:Species: Agkistrodon contortrix contortrix (southern copperhead)
 C>Date: 30-Sep-1992 #sequence_revision 09-Apr-1998 #text_change 18-Jun-1999
 C:Accession: S66259; A41827; S25461; A37303
 R:Sellistre de Araujo, H.S.; Ownby, C.L.
 Arch. Biochem. Biophys. 320, 141-148, 1995
 A:Title: Molecular cloning and sequence analysis of cDNAs for metalloproteinases from brown
 A:Reference number: S66259; MUID:95314311; PMID:7793974
 A:Accession: S66259
 A:Molecule type: mRNA
 A:Residues: 1-411 <SEL>
 A:Cross-references: EMBL:U18233; NID:g603214; PIDN:AAC59703.1; PID:g603215
 R:Randolph, A.; Chamberlain, S.H.; Chu, H.L.C.; Retzius, A.D.; Markland Jr., F.S.; Masia
 protein Sci. 1, 590-600, 1992
 A:Title: Amino acid sequence of fibrolase, a direct-acting fibrinolytic enzyme from Agki
 A:Reference number: A37303; MUID:93278288; PMID:1304358
 A:Accession: A41827
 A:Molecule type: protein
 A:Residues: 191-192,'R',194-199,'Q',201-240,'Q',242,'T',244-357,'AAM',361-381,'T',383-399
 A>Note: 379-Glu and 382-Leu were also found
 A:Note: a variant lacking 192-Gln was also found
 R:Guan, A.L.; Retzius, A.D.; Henderson, G.N.; Markland Jr., F.S.
 Arch. Biochem. Biophys. 289, 197-207, 1991
 A:Title: Purification and characterization of a fibrinolytic enzyme from venom of the sc
 A:Reference number: S17490; MUID:91378546; PMID:1898066
 A:Accession: S25461
 A:Molecule type: protein
 A:Residues: 331-341 <GUA>
 C:Genetics:
 C:Gene: ACLPREF
 C:Superfamily: atrolysin C
 C:Keywords: hydrolase; metalloproteinase; pyroglutamic acid; venom; zinc
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-190/Domain: propeptide #status predicted <PRO>
 F:191-411/Product: fibrolase status experimental <WAY>
 F:191/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime
 F:308-388/Disulfide bonds: #status experimental
 F:333,337,343/Binding site: zinc (His) #status predicted
 F:334/Active site: Glu #status predicted
 F:348-355,350-372/Disulfide bonds: #status predicted

Query Match 9.78; Score 307; DB 1; Length 411;
 Best Local Similarity 26.4%; Pred. No. 1.6e-15;
 Matches 112; Conservative
 Indels 66; Gaps 17;

Qy 24 EWVPIRLDPDINGRRYYWRGPESDGGLFIQTAFQEDFYHLTLTDAQFLAPAFSTE 83
 :
 Db 32 EVVIYPKVTVPVRG-----AVQPYEDAMQYEKVNGEPPVVLHLEKNGLFSEDSYSETH 85
 Qy 84 LGVPLQGLTGSSDLRCRCFYGDVNAPEDPSFAVSLCGILRGATGYCARGAEVVISPLPNAS 143
 :
 Db 86 YSPDGRETTIPLYVEDHCYYHGRTEINDASTASINACNLKGHKLOEMYLEIPL----- 141
 Qy 144 APAORNISOGAHLLOR-RGVPGGSGDPDTSRCGVASGNWNPAIRLDRALDPYKPRRAGGESR 202

F:296/Active site: Glu #status predicted

```
Query Match          9.7%; Score 306; DB 2; Length 571;
Best Local Similarity 23.4%; Pred. No. 2.9e-15;
Matches 139; Conservative 76; Mismatches 217; Indels 162; Gaps 30;

QY 12 AGRTAGGPEPREVVPIRLDPDINGRYYWRGPDSDGGLIFQITAFQEDF-YLHLTP 70
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 ATRPKGAVQPKYEDA--MOYEFKVNGEPPVVLHLEKNG-----LFSKDYSEIHSP 49

QY 71 DAQFLAPAFSTEHLGVLPGLOGTLGSSDLRCFCYSGDVNAEPDSFAAIVSLCGGLRGAGYR 130
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 50 DGREITTPYPPVED-----HCYHGRINDADASTASISACNGLKGYEKLQ 93

QY 131 GAEVYISP--LPNASAPAAQRNSOGAHLQRRGVPGPSDPTSRGCVASGWN----- 181
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 94 RETYFIEPLKLPDSEAH-----VKYENVE--KEDEAPKMGCVTONKSYEPIKK 142

QY 182 -----PALRALDPYKPRRAGFGESESRSSRAKRFVSIPIRYVELLVVADESMVKFHG 235
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 143 ASQLAFTAEQQOQYDPYK-----YIEFFVVVDQGTVTNN 176

QY 236 ADLEHYLLTLTAARLYRHSILNPI-NIVVVKVLL--LRDRDSGPKVT--GNAALTLR 290
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 177 GDLD-----KIKARWELANIVNEIFRILYMEVALVGLTEWSNGDKITVKPDVDYTLN 229

QY 291 NFAQWQKLNKVSOKHPEYWDTAILFTRODLCCGATTCDTLGMADVMTCDPKRSCSVIED 350
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 230 SFAEWRKTDLLLRKKH-----DNAQLLTAIDFNG-----PTIGYAYIGSMCHPKRSVGIVQD 281

QY 351 ---DGLPSAFTHAELGHVFMNPHDNVYKVEEFGKLRANH---MMSPTLIQIDRANPWS 404
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 282 YSPINLVVAVTMAHEMGNHGIHDT-----GSCSGDYPCIMGPT--ISNEPSKFFS 332

QY 405 ACSAAITDFLDSHGGLCLDQD-----SKPISLPELPGASTTISOCELAFAVGSKP 458
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 333 NCSYIQWDFIMNHNPICINEPLGTDIISPPVCGNELL-----EVGECD----- 378

QY 459 CPYMQYCTKWCTGKAGQMVQTRHFPWADGTCGSGGKCLKLGACVERHNLNKRVD-- 516
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 379 CGTPENCNECCDA-----ATCKLK-----SGSQCHGD-----CCEQCKFSKSGTECR 422

QY 517 GSWAKWDPYGPCC---SRTCGGVQLARROCTNTPANGKGYC--EGVRVKYRSC 565
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 423 ASMSECDPAEHCIGQSECCADVFHKNQ---PCLDNYG-YCYNGNCPIMYHQC 472
```

RESULT 13

S66260
metalloproteinase (EC 3.4.24.-) precursor - southern copperhead
C:Species: Agkistrodon contortrix contortrix (southern copperhead)
C:Date: 17-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 22-Jun-1999
C:Accession: S66260; S74263
R:Selliste de Araujo, H.S.; Ownby, C.L.
Arch. Biochem. Biophys. 320, 141-148, 1995
A:Title: Molecular cloning and sequence analysis of cDNAs for metalloproteinases from b
A:Reference number: S66259; MUID:95314311; PMID:7793974
A:Accession: S66260
A:Molecule type: mRNA
A:Residues: 1-407 <SEL>
A:Cross-references: EMBL:U18234; NID:g603216; PIDN:AAC59704.1; PID:g603217
A:Accession: S74263
A:Molecule type: protein
A:Residues: 188-206 <SEA>
C:Superfamily: atrolysin C
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-187/Domain: propeptide #status predicted <PRO>
F:188-407/Product: metalloproteinase #status experimental <MAT>
F:304-384, 344-351, 346-368/Disulfide bonds: #status predicted
F:329,333,339/Binding site: zinc, catalytic (His) #status predicted
F:330/Active site: Glu #status predicted

```
Query Match          9.6%; Score 304; DB 2; Length 407;
Best Local Similarity 25.3%; Pred. No. 2.7e-15;
Matches 110; Conservative 69; Mismatches 164; Indels 92; Gaps 18;

QY 24 EVVPIRLDPDINGRYYWRGPDSDGGLIFQITAFQEDFVLHLTPDAQFLAPAFSTEH 83
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 32 EVVYPRKVTVPVPGK-----AVQPKYEDAMQYELKYNGEPPVVLHLERNKGLFSKDYSETH 85

QY 84 LGVPLGLOGTLGSSDLRCFCYSGDVNAEPDSFAAIVSLCGGLRGAGYRGAEYVISPLPNAS 143
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 86 YSPDGRKITTYPPVEDHCYHGRINDADASTASISACNGLKGFKLQGEWYLIEPLELSD 145

QY 144 APA-----AQRNSOGAHLQRRGVPGPSDPTSRGCVASGWNPAILRALDPVKPRR 195
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 146 SEAHAVFKYENVEKEDEAPKI-----CGVTQNW-----ESTEP-- 178

QY 196 AGFGESESRSSRAKRFVSIPIRYVELLVVADESM-VKFGAD--LEHYLLTLATAARL 252
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 179 -----IKKASQLNLNYQRYVELVTVVDHGMVTKYNGSDSKIROWVHQWMTMKES 230

QY 253 YRHPSTLNPINIVVVKVLLLRDR-DSGPKVTGNAALTIRNFCAWOKK--LNKVSOKHPEY 309
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 231 YRYWYI--DISLAGVEIWSNKDLIDVQPA-----ARHTLDSFGWEERDILLHRISH-- 279

QY 310 WDTAILFTRODLCCGATTCDTLGMADVMTCDPKRSCSVIEDDGLPS---AFTTAHELGHV 366
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 280 -DNAQLLTSTDFG-----PTIGLAYVTGMCDDPKLSTGVVEDHKSINFLVAVTMAHEMGNH 334

QY 367 FMNPHDNVYKVEEFGKLR-----NHMSPTLIQIDRANPWSACSAIITDFLDSHGGLCL 423
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 335 LGMRHDT-----GSCSGGYSCIMSP-VISDDSPKYFSNCVSIQCWDFIMKENPOCI 385

QY 424 LQDP-----SKPIS 432
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 386 LNKRLTDTVSTPVS 400
```

RESULT 14

S55270
catrocollastatin precursor - western diamondback rattlesnake
C:Species: Crotalus atrox (western diamondback rattlesnake)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C:Accession: S55270
R:Zhou, Q.; Smith, J.B.; Grossman, M.H.
Biochem. J. 307, 411-417, 1995
A:Title: Molecular cloning and expression of catrocollastatin, a snake-venom protein
A:Reference number: S55264; MUID:95251603; PMID:7733877
A:Accession: S55270
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-609 <ZHO>
A:Cross-references: GB:U21003; NID:g710353; PIDN:AAC59672.1; PID:g710354
C:Superfamily: mouse meltrin alpha, disintegrin homology

Query Match

```
Query Match          9.6%; Score 302.5; DB 2; Length 609;
Best Local Similarity 23.5%; Pred. No. 5.8e-15;
Matches 129; Conservative 80; Mismatches 218; Indels 121; Gaps 25;

QY 50 DOGLIFQITAFQEDFVLHLTPDAQFLAPAFSTEHLGVLPGLOGTLGSSDLRCFCYSGDVNA 109
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 52 EDAMQYELKYNGEPPVVLHLGNKNGKLFSDYSETHYSPDGRREITYYPLVEDHCYHGRIN 111

QY 110 EPDSFAAVSLCGGLRGAGYRGAEYVISP--LPNASAPAAQRNSOGAHLQRRGVPGGPS 167
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 112 DADSTASISACNGLKGFKLQGEWYLIEPLKLPDSEAHAVKYENVE-----KE 160

QY 168 GDTSERCVSGSWNPAILRALDPYKPRRAGFGESESRSSRAKRFVSI-----RY 219
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161 DEALKMCGVTQNW-----ESYEPIK-----KASQLVVTAEHQKYNPFRR 199

QY 220 VETLVVADESMVKFHGADLEHY---LLTLLATAARLYRHPSTLNPINIVVVKVLLLRD 276
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 VELFLVVDKAMWTKNNGDLKIKRMYEIVTVNVEIYIMY----IHVALVGLWEIWSND 255
```

265	Db	NFLSWREQ---NLOQHPH---DNYOLITGVDFIGS-----TVGLAKVSALCS-RHSGAVNOD 315
351	QY	--DGLFSATTTAHELGHVENPHD-NVKVC-----EEVFGKLRANHMS--PTLIQIDR 399
316	Db	HSKNSIGVASTMAHELGHNLGMSHDEIPCCYCPEPREGGGCIIWTESIGSKFPRI-----370
400	QY	ANPWSACSAAIIITDFLDLSHGHDCLDQP-----SKPI-----SLPDLPG 439
371	Db	--FSRCKIDLESFVTKPQTGCLTNVPDYNRVFGVPGVGNLFVHEHGEQDCGTPQDCQN 427
440	QY	-----ASYTLSSQOCEL-----AFG 453
428	Db	PCCNATTQLVKGAECASGTCHECKKPKAVECRLSKDKDLEEFCDGKPKTCPEDAFQ 487
454	QY	VGSKPCP-----YMOYCTKLW-----CTGKAK-----G 476
488	Db	QNGTPCPGGYCFDGSCTPLAQQRDLUMGPGARVAADSCYFTSIPPCNGRMYSGRINRCG 547
477	QY	QMVQOTRHPFW-----ADTSGGEGKLCILKGACVER 507
548	Db	ALYCEGGGKPLERSECTFSSNHGVCHALGTGSDNIDTFELVLQGTCEEKGKVCMDGSC---604
508	QY	HNLNKHRYDGSWAKWDYPGCSRTCGGVQJLARRQC 543
605	Db	ODLWVRSENCASACNNHGVCH-----KREC 631

Search completed: April 11
Job time : 17.4848 secs

```
Search completed: April 29, 2003, 17:13:25
Job time : 17.4848 secs
```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 7.18023 Seconds
(without alignments)
3367.676 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_583

Perfect score: 3163

Sequence: 1 MLLGLILTAPAGTAGGFE.....SCNLEPCPSSASGKSFREQ 583

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1562.5	49.4	968	1 ATSL_MOUSE	P97857 mus musculus
2	1560.5	49.3	967	1 ATSL_HUMAN	Q9uh18 homo sapien
3	1547	48.9	967	1 ATSL_RAT	Q9uwl1 rattus norv
4	1362	43.1	890	1 ATSL_HUMAN	Q9up79 homo sapien
5	1338.5	42.3	837	1 ATSL_HUMAN	Q75173 homo sapien
6	1329.5	42.0	905	1 ATSL_MOUSE	P57110 mus musculus
7	1242	39.3	930	1 ATSL_MOUSE	Q9r001 mus musculus
8	1232.5	39.0	1629	1 ATSL_HUMAN	Q9p2n4 homo sapien
9	1212.5	38.3	930	1 ATSL_HUMAN	Q9una0 homo sapien
10	1168.5	36.9	930	1 ATSL_RAT	Q9esp7 rattus norv
11	937.5	28.6	997	1 ATSL_HUMAN	Q9ukp4 homo sapien
12	891.5	28.2	1077	1 ATSL_HUMAN	Q9h324 homo sapien
13	876.5	27.7	1593	1 ATSL_HUMAN	P58397 homo sapien
14	788.5	24.9	1211	1 ATSL_HUMAN	O95450 h adamts-2
15	775	24.5	1205	1 ATSL_BOVIN	P79331 b adamts-2
16	752	23.8	1205	1 ATSL_HUMAN	O15072 homo sapien
17	725.5	22.9	860	1 ATSL_HUMAN	Q9ukp5 homo sapien
18	599	18.9	207	1 ATSL_BOVIN	Q9tt92 bos taurus
19	599	18.9	245	1 ATSL_BOVIN	Q8tt93 bos taurus
20	351	11.1	824	1 AD08_HUMAN	P78325 homo sapien
21	348.5	11.0	776	1 AD28_MACEFA	Q9xsl6 macaca fasc
22	329	10.4	903	1 AD12_MOUSE	Q61824 mus musculus
23	326	10.3	813	1 AD33_HUMAN	Q6bz11 homo sapien
24	324	10.2	774	1 AD28_MOUSE	Q9jln6 mus musculus
25	323.5	10.2	775	1 AD28_HUMAN	Q9ukg2 homo sapien
26	314	9.9	909	1 AD12_HUMAN	O43184 homo sapien
27	306	9.7	571	1 DISJ_BOVIFA	P30431 bothrops ja
28	301	9.5	826	1 AD08_MOUSE	Q95910 mus musculus
29	300	9.5	956	1 AD19_HUMAN	Q9h013 homo sapien
30	297	9.4	480	1 DISA_TRIGA	P15503 trimeresuru
31	296	9.4	413	1 ACIA_AGRAC	Q9pw35 agkistrodon
32	295.5	9.3	920	1 AD19_MOUSE	Q95674 mus musculus
33	284	9.0	414	1 HRTD_CROAT	P15167 crotalus at

RESULT 1

ID	ATSL_MOUSE	STANDARD;	PRT;	968 AA.
AC	P97857: O54768;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).			
DE	ADAMTS1.			
GN	Mus musculus (Mouse).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OC	NCBI_TaxID=10090;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=129/SvJ;			
PC	MEDLINE=98110583; PubMed=9441751;			
RX	Kuno K., Lizasa H., Ohno S., Matsushima K.;			
RA	"The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";			
RT	Genomics 46:466-471(1997).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=97150761; PubMed=8995297;			
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F., Matsushima K.;			
RC	"Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene.";			
RT	J. Biol. Chem. 272:556-562(1997).			
RL	[3]			
RN	CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.			
RP	MEDLINE=99303657; PubMed=10373500;			
RA	Kuno K., Terashima Y., Matsushima K.;			
RC	"ADAMTS-1 is an active metalloproteinase associated with the extracellular matrix.";			
RT	J. Biol. Chem. 274:18821-18826(1999).			
RL	[4]			
RN	FUNCTION.			
RP	MEDLINE=203895568; PubMed=10930576;			
RA	Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M., Ohno H., Matsushima K.;			
RC	"ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";			
RT	FEBS Lett. 478:241-245(2000).			
RL	[5]			
RN	FUNCTION, AND INDUCTION.			
RP	MEDLINE=20243757; PubMed=10781075;			
RA	Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W., Richards J.S.;			
RC	"Progesterone-regulated genes in the ovulation process: ADAMTS-1 and cathepsin L proteases.";			
RT	Proc Natl Acad Sci U.S.A. 97:4689-4694(2000).			
RL	CC			
CC	-1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH			

Q9rlv6 mus musculus
Q63180 rattus norv
Q9p0k1 homo sapien
Q35227 mus musculus
P34182 crotalus at
Q923w9 mus musculus
Q13444 homo sapien
Q9h2u9 homo sapien
Q28475 macaca fasc
P30403 agkistrodon
Q42596 xenopus lae
Q75078 homo sapien

ALIGNMENTS

CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-1-LEU-1692
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX.
CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC CELLS OF PREOVULATORY FOLLICLES.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB001735; BAA24501.1; ALT_INIT.
DR EMBL; D67076; BAA11088.1; ALT_FRAME.
DR MEROPS; M12.222; .
DR MGD; MGI:109249; Adamts1.
DR InterPro; IP001762; Disintegrin.
DR InterPro; IP002870; Pep_M12B_propep.
DR InterPro; IP001390; Reprolysin.
DR InterPro; IP000884; TSP1.
DR InterPro; IP000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00092; TSP1; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Repeat; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 48
FT PROPEP 49 253
FT CHAIN 254 968
FT SITE 206 206
FT METAL 402 402
FT ACT_SITE 403 403
FT METAL 406 406
FT METAL 412 412
FT DOMAIN 477 559
FT DOMAIN 560 617
FT DOMAIN 618 725
FT DOMAIN 726 850
FT DOMAIN 851 909
FT DOMAIN 910 968
FT DOMAIN 195 199
FT POLY-ARG.
FT CARBOHYD 721 721
FT CARBOHYD 721 721
FT CARBOHYD 765 765
FT CARBOHYD 783 783
FT CARBOHYD 946 946
FT MUTAGEN 403 403
FT CONFLICT 335 335
FT CONFLICT 425 425
FT SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;

Query Match 49.4%; Score 1562.5; DB 1; Length 968;
Best Local Similarity 50.4%; Pred. No. 3.2e-112;
Matches 315; Conservative 79; Mismatches 154; Indels 77; Gaps 16;
QY 1 MLLGILTLAAGTAGG--FEPEREVVVPIRLDPDINGRRYYWRGP-EDSGDGLLFOI 57
DB 37 LLLASITMLLCARGAHGRPTDEELVLP-SLE-----RAPGHDSITTRL--RL 83
QY 58 TAFQDEFLHTLPDPAFLAPAFSTEHGLGVPLQGLTGG-----SDLRRCFFSGDVNA 109
DB 84 DAFGQLLKLQPSGGFLAPGFTLTQV-----GRSPGEAQLDPTGDLAHCFFSGVNG 138
QY 110 EPDSPAAYSLCGLRGAFYRGAEYVISPFPNAS-----APAAQRNSQGA-----HLLQRRG 161
DB 139 DPGSAALSLCEGVRGAFYLGEEFFIQAPGVATERLAPAVPEESARPOFHILRRR- 197
QY 162 VPGSPGDPTRSCGVASGWNFAILRALDPYKPRRAGGESRRR----- 206
DB 198 ----RRSGGAKCGYMD-----DETLPSTDSRPSQNTNRNQPVRDPTPDAGKP 243
QY 207 ---SGRAKRVSPRYVETLWADSMVKFPGADLEHYLLTLLAARLYRHPISILNPI 262
DB 244 SGPISIRKRVSPRYVETLWADSMVKFPGADLEHYLLTLLAARLYRHPISILNPI 303
QY 263 NIVYKVLRLDRDGPVKVTGNAALTILNFCAMQKLNKYSKHPYWDYTAILEFTRODLC 322
DB 304 SLVYKVLIVYEEQKPEVTSNAALTILNFCAMQKLNKYSKHPYWDYTAILEFTRODLC 363
QY 323 GATTCDDTIGMADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDKNVKEEVFG 382
DB 364 GSHTCDDTIGMADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDKNVKEEVFG 423
QY 383 KLRANHMSPTLIQIDRANPWSAGSAIITDLDGSHGDCLLDQPSKIPSLPDLPGASY 442
DB 424 VTGDSHLNMLSLDHSQWSPSCAYVMTSFLDNGHGECLMDKPNQPIKLPSDLPGTLY 483
QY 443 TLSQCELAFGVGSKPCP-YMQYCTKLWCTGKAGQVMQVQTRHFPFWADGTCGSGEGKCLCK 501
DB 484 DANROCFTFGEESKHCPDAASTCTTLWCTSGLLVCQTKHFPWADGTCGSGEGKCVCS 543
QY 502 GACVERHNLNKH---RVDGSAWADPYPGCPSCRTCGGVQVLAARROCTNPTPANGKYGCV 558
DB 544 GCVNKTDM-KHATPVHSGWGPWGDGCSRTCGGVQVYTMRECDNFPVKNKGKCYCEGK 602
QY 559 RVKRSCHNLEPCSSASGKSFREEQ 583
DB 603 RVKRSCHNLEPCSSASGKSFREEQ 626
RESULT 2
ATSL_HUMAN
ID ATSL_HUMAN STANDARD; PRT; 967 AA.
AC Q9UHI8; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
GN ADAMTS1 OR METH1 OR KIAA1346.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
FT "Cloning, characterization and mapping on human chromosome 21 of the
FT orthologue of murine Adams-1.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC TISSUE=Heart;
RX MEDLINE=99367466; PubMed=10438512;

[illegible]


```
Db 397 IEDDGLHAAFTVAHEIGHLLGLSHDDSKFCENFOTTDKRLMSILLTSIDASKPWSKCT 456
QY 408 AAILTFDLSGHCCLLPQSPKPSLIPEDLPASYSYTLISQOCCLAFGVSGKPCPYMQYCTK 467
Db 457 SATITFEDDGHGNCCLLPKQILGLPELPGQYDATQOCNLTFGPESYVCPGMDYCAR 516
QY 468 LWCTGKAGQVMQVQTRHFPWADGTSCEGKCLKLAGACVER-----HNLNKHVRVDGNAKW 522
Db 517 LWCAVVRQGMVCLTKKLPVAVEGPCGKGRVCLQKGVKTKKKYKYSTSSH---GNNGSW 573
QY 523 DPYGCSTCCGGVQLARRQCTNTPANGGYKCGVRVYKSCNLEPCPSASGKSPREE 582
Db 574 GPMQCSRSKGGVQFAYRHCNCPAPRNSGRYCTGKRAIYRSCSVTPCP--PNGKSPRHE 631
QY 583 Q 583
Db 632 Q 632

RESULT 8
AT99_HUMAN
AC Q9P2N4; Q9NR29; STANDARD; PRT; 1629 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
GN ADAMTS9 OR KIAA1312.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A. (SHORT ISOFORM).
RT TISSUE=Fetal.
RX MEDLINE=20396138; PubMed=10936055;
RA Clark M.E., Kellner G.S., Turbeville L.A., Boyer A., Arden K.A.,
RA "Adamts 9, a novel member of the ADAM-TS/Metallospodin gene
RT family";
RL Genomics 67:343-350(2000).
RN [2]
RS SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirotsawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RL for large proteins in vitro.";
RN DNA Res. 7:65-73(2000).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.
CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LONG,
CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN
CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR
CC THYMUS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF261918; AAF89106.1; -
CC EMBL: AB037733; BAA92550.1; -
CC HSP: P15167; IATL.
CC MEROPS: M12.021; -.
CC Genew: HGNC:13202; ADAMTS9.
CC MIM: 605421; -.
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Repolysin.
CC InterPro: IPR000884; TSP1.
CC InterPro: IPR000130; Zn_MTpeptdse.
CC Pfam: PF00090; tsp_1; 11.
CC Pfam: PF01421; Repolysin; 1.
CC Pfam: PF01562; Pep_M12B_propep; 1.
CC SMART: SM00209; TSP1; 12.
CC PROSITE: PS02015; ADAM_MEPRO; 1.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE: PS50092; TSP1; 9.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 18
FT PROPEP 19 287
FT CHAIN 288 1629 ADAMTS-9.
FT DOMAIN 509 587 DISINTEGRIN-LIKE.
FT DOMAIN 589 642 TSP TYPE-1 1.
FT DOMAIN 645 752 CYS-RICH.
FT DOMAIN 753 880 SPACER.
FT DOMAIN 999 1053 TSP TYPE-1 2.
FT DOMAIN 1056 1108 TSP TYPE-1 3.
FT DOMAIN 1111 1156 TSP TYPE-1 4.
FT DOMAIN 1184 1239 TSP TYPE-1 5.
FT DOMAIN 1240 1295 TSP TYPE-1 6.
FT DOMAIN 1322 1383 TSP TYPE-1 7.
FT DOMAIN 1386 1439 TSP TYPE-1 8.
FT DOMAIN 1445 1498 TSP TYPE-1 9.
FT DOMAIN 1501 1554 TSP TYPE-1 10.
FT DOMAIN 1562 1612 TSP TYPE-1 11.
FT SITE 223 223 CYSTEINE SWITCH (POTENTIAL).
FT METAL 434 434 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 435 435 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 438 438 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 444 444 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 749 749 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 840 840 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1213 1213 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1267 1267 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPPLIC 1064 1072 CLVTCGKGH -> VRWECYFP (IN
FT SHORT ISOFORM).
FT VARSPPLIC 1073 1073 MISSING (IN SHORT ISOFORM).
FT CONFLICT 367 367 F -> L (IN REF. 1).
SQ SEQUENCE 1629 AA; 182649 MW; C1C4CEFF58B941F CRC64;
Query Match 39.0%; Score 1232.5; DB 1; Length 1629;
Best Local Similarity 40.6%; Pred. No. 1.3e-86;
Matches 252; Conservative 87; Mismatches 210; Indels 71; Gaps 13;
QY 22 EREVVVPIRLD-----DINGRRYYW----RGPEDSGDQGLIFQITAFQ 61
Db 47 EYEIVSPIRVNALGPEFFTNVHFKTRTSINSATDPWAFASSSSSTSSQAHRLSAGF 106
QY 62 EDFYHLHPDQAFAPASTHEHLGVP-----LQLTGGSSDLRRRCFYSGDVNAEPDPAFV 117
Db 107 QQFLNLTANAGFIAPLFTVTLGLTGPVGNQTKFYSEEAELKHCYKGYVNTNHTAVI 166
```


entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF140675; AAD56358.1; -
CC HSP; P15167; 1ATL.
CC MEROPS; M12.231; -
CC Genew; HGNC:223; ADAMTS7.
CC MIN; 605009; -
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; zn_Mtpeptdse.
CC Pfam; PF00090; tsp_1; 1.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSP1; 2.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS0215; ADAM_MEPRO; 1.
CC PROSITE; PS0092; TSP1; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
KW SIGNAL; 1 27 POTENTIAL.
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 BY SIMILARITY.
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;
Query Match 29.6%; Score 937.5; DB 1; Length 997;
Best Local Similarity 39.4%; Pred. No. 3.2e-64;
Matches 250; Conservative 69; Mismatches 216; Indels 99; Gaps 29;
QY 1 MLLILGILTLAFAGRTAGGEPER---EVVVPRLDP-----DINGRR- 39
Db 18 LLLICALAPGAPGAPGRATEGRAALDIVHVPVRVDAGGSFSLYELWPRALRRKRVSVRRD 77
QY 40 ---YY---WRGPEDSGDQGLIQTAFQEDFYHLTPDAQFLAPAF---STEHLGVPLQGL 91
Db 78 APAFYELQYRGRE-----LRFNLTAQ-----HL-----LAPGVSETRRG---GL 116
QY 92 TGGSSDLR---RCFYSGDVNAEPD---SFAAVSLCGGLRGAGFYRGAEYVISPFPNASH 144
Db 117 --GRAHRAHTFACHLLGEVQ-DPELEGLAIAISACDGLGVFQLSNEDYFIEPLD--SA 171
QY 145 PAAQRNSQGAHLQRRGVP-----GGPSGDPTSRGCVASGWNPAILLRALDPKYPKRAFG 199
Db 172 PARPGHAQ-PHVYKRAQAPERLAQRGDSSAP-STCGV-----QVYELESRRERW 219
QY 200 ESRSRRSRGRKRF-----VSIPRYVETLVVADESVMVKPHG-ADLEHLLTLTAARLYR 254
Db 220 EQRQWRRLRLRLHRSVSKEKVEVTLVVAADAKVVEYHQVQVSVYLTIMNVAAGLFH 279
QY 255 HPSILPINIVVVKVLLLRDRSGPKVTGNAALTNRNCAWQKLNKVKSDKHPYWDTAI 314
Db 280 DPSIGPHITIVRVLLLEDEEDLKITHADNTLKSFCWKWSINMKGDAHPHHDTAI 339
QY 315 LFTRODLGATT--CDTLGMADVGMCDPKRCSVIEDDGLPSAFTTAHELGHVFNMPHD 372
Db 340 LLTRKDLCAAMNRPCTETGLSHVAGMCQPHRCSINEDTGLPLAFTVAHELGHSEGIQHD 399

Query Match 36.9%; Score 1168.5; DB 1; Length 630;
Best Local Similarity 54.9%; Pred. No. 3.4e-82;
Matches 208; Conservative 52; Mismatches 114; Indels 5; Gaps 4;
QY 209 RAKREFVSIPRYVETLVVADESVMVKPHGADLEHLLTLTAARLYRHPHSILNPIVNVVK 268
Db 2 RTRKFAISLRFVETLVVADDKMAAFHGAGLKHVLLTVMAAAAKAFKPSIRNPVNLVTR 61
QY 269 VLLLRDRDSQPKVTGNAALTNRNCAWQKLNKVKSDKHPYWDTAI LFTRODLGATTCD 328
Db 62 LVILGSGQEVQPGVSAQAQTLRFCTWQKLNPNPNDSDPHEDTAI LFTRODLGVTSCD 121
QY 329 TLGMADVGMCDPKRCSVIEDDGLPSAFTTAHELGHVFNMPHDNPKVCEEVFGK-LRAN 387
Db 122 ALGMAGVTGCDPARSAIVEDDGLQSAFTAHELGHVFNMLHNSKPCANLNGQSSSR 181
QY 388 HMSPTLIQIDRANPWSACSAIITDLSGHGDCILLDPQSPKISIPEDLPASVYTLSSQ 447
Db 182 HVNAPVMAHVDPEEPWSPCSAREITDPLNGYGHCLLDKPEAPLHLPVTFPGKDYDADRQ 241
QY 448 CELAFGVGSKPCVM-QYCTKLACTGKAKGMVQCOTRHPHWDGTS CGEGKLCCKGACVE 506
Db 242 COLTFGPDSSHCQLPPLPCCAALMCFGLHNGAMCQTAHSPWADGTCGGAQAQCMGRCLH 301
QY 507 RHNLNKRHVD--GSWAKWDYGPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVRYKYS 564
Db 302 VDQLKDFNIPQAGGWGPMGWDGSRCTCGGGVQFSSRDCTKPVPRNGKYCEGRRTPERS 361
QY 565 CNLEPCPSSASGKSFREEQ 583
Db 362 CNTKNCP-HGSALTFFREEQ 379
RESULT 11
AT57_HUMAN STANDARD; .PRT; 997 AA.
AC Q3UKP4;
CD MEDLINE=9395124; PubMed=10464288;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
GN ADAMTS7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=9395124; PubMed=10464288;
RX Hurskainen T.L., Hirohata S., Seidlin M.F., Apte S.S.;
RA "ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of
RT Zinc Metalloproteases";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,
CC LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

FT	DOMAIN	947	995	TSP TYPE-1 4.
FT	DOMAIN	996	1315	SPACER 2.
FT	DOMAIN	1316	1364	TSP TYPE-1 5.
FT	DOMAIN	1367	1423	TSP TYPE-1 6.
FT	DOMAIN	1426	1471	TSP TYPE-1 7.
FT	DOMAIN	1426	1471	TSP TYPE-1 8.
FT	DOMAIN	302	305	POLY-GLU.
FT	SITE	208	208	CYSTEINE SWITCH (POTENTIAL).
FT	METAL	392	392	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	393	393	BY SIMILARITY
FT	METAL	396	396	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	402	402	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	105	105	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	215	215	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	685	685	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	951	951	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1104	1104	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1275	1275	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1300	1300	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1320	1320	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1371	1371	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1503	1503	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	1593 AA;	177545 MW;	07F9F48E63BD83A3 CRC64;
QY	Query Match	27.7%;	Score 876.5;	DB 1; Length 1593;
Db	Best Local Similarity	36.6%;	Pred. No. 2.7e-59;	
QY	Matches 21%;	Conservative	92;	Mismatches 230;
Db				Indels 49;
QY	21 PEREVVVPRLD-----	PDINGRRYYRWGPEDSGDGLIFQITAFQEDFYHLH 68		
Db	50 PEYHVGVPRVDASGHFLSYGLHYPTTSRR--	KRDLDSGDWVYRISHEEKDLFFNL 106		
QY	69 TPDAQFLAPAFSTHEHLGVPLQGLTGSSDLRRCFCYSGDYNABPD--	SFAAVSLCGLGRLCA 126		
Db	107 TVNQGFGLNSYIMEKRYGNLUSHYKMASSAPICHLSGTVLQOQTRVGTAAALSACHGLTGF 166			
QY	127 FGYRGAEYVISPFPNAPAAQRNSQGAHLQRRGYPGGPGSDPTSRGCVAGSNPAILR 186			
Db	167 FQLPHGDFTEPV--KKHPLVEGGYHPHIVYRQKVP--ETKEPT--	CGLKDSVNTSQK 220		
QY	187 ALDPYKPRRACFGESRRSRSGRAKRFVSPTRPVETLVVADESVMKFKHAD--LEHYLLTL 245			
Db	221 ELWREKWERNL--PSKSLSR-----	RSISKERWETLVVADTKMIYHGSENVESVILII 274		
QY	246 LATAARLYRHPSTLININIVVVKLLLRDRSDSPKVTGNAALTFRNFCQAKKLKLVSKD 305			
Db	275 MNVVTGLFHPNSIGNAHIVVRLILLEEEQGLKIVHHAETKLSFCFKWKSINPKSDL 334			
QY	306 HPEYWDTAFLTTRQDLGGA--TTCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTAHEL 363			
Db	335 NPVHHVDVAVLLTRKIDICAGENRPECETLGLSHLSCMQPHRSCNINEDSGLPPLAFTTAHEL 394			
QY	364 GHVFNMPHDNVK--VCEVEFCKLRANHMMSPTLIQIDRANPWSACSAAITDFLDSHGDC 422			
Db	395 GHSEGIQHDGKENDCEPVG---RHPYIMSRQLQDYDTPLTWKSCSEYITRFLDRGWGFC 451			
QY	423 LLDPQSKP--ISLPEDLPGASYTLSSQOCELAFGYGVSKPCPYMQ--YCTKLMTCTGAKAKQMV 480			
Db	452 LDDIPKKGLKSKVIAPGVYIDVHHQQLQYGPNAFCQEVENVQTLWCSVKG----FC 507			
QY	481 QTRHFPWADGTSQEGKLCIKGACVERHNLNK--HRVDGSAWKWDYVGCPSRTC GGSGVOL 538			
Db	508 RSKZDAADGTCQEGKKWCWAGKCI---TVGKKPESIPGGWGWSWPSHSCSRCTCGAGVOS 564			
QY	539 ARROCTNPTPANGKCYGEVRYRSCNLCRPPCPSSAGSKSFREQ 583			
Db	565 AERLCNNPEPKFGKYCTGERKRYRLCNVHPCRSEA--PTFRMQQ 607			

RESULT 14

AT22_HUMAN STANDARD; PRT; 1211 AA.
 AC O95450;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase with thrombospondin motifs 2) (ADAM-TS2)
 DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI)
 DE (Procollagen I/II amino-propeptide processing enzyme).
 GN ADAMTS2 OR PCNP1 OR PCPNI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LPNP1 AND SPNP1), AND DISEASE.
 RC TISSUE-Skin;
 RX MEDLINE=99347935; PubMed=10417273;
 RA Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E., Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W., Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.;
 RA "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene.";
 RL Am. J. Hum. Genet. 65:308-317(1999).
 CC -!- FUNCTION: Cleaves the propeptides of type I and II collagen prior to fibril assembly. Does not act on type III collagen. May also play a role in development that is independent of its role in collagen biosynthesis.
 CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains at Ala-1-Gln.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO COLLAGEN TYPE XIV (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNP1 (SHOWN HERE) AND SPNP1; ARE PRODUCED BY ALTERNATIVE SPLICING. SPNP1 HAS NO SIGNIFICANT N-PROCOLLAGEN PEPTIDASE ACTIVITY.
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).
 CC -!- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos syndrome type VIIC (EDS-VIIC), a recessively inherited connective-tissue disorder characterized clinically by severe skin fragility and joint hypermobility and biochemically by the presence in skin of procollagen incompletely processed at the N-terminus.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ003125; CAA05880.1;
 CC MEROPS; M12.301;
 CC Genew; HGNC:218; ADAMTS2.
 CC MIM; 604539;
 CC MIM; 225410;
 CC InterPro; IPR001762; Disintegrin.
 CC InterPro; IPR002870; Pep_M12B_propep.

DR InterPro; IPR001590; Reprolysin.
 DR InterPro; IPR00884; TSP1.
 DR InterPro; IPR000130; Zn_MTPeptidse.
 DR Pfam; PF00090; tsp_1; 4.
 DR Pfam; PF01421; Reprolysin; 1.
 DR Pfam; PF01562; Pep_M12B_propep; 1.
 DR SMART; SM00209; TSP1; 4.
 DR PROSITE; PS00215; ADAM_MEPRO; 1.
 DR PROSITE; PS00092; TSP1; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
 KW Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
 KW Alternative splicing; Ehlers-Danlos syndrome.
 FT SIGNAL 1 29 POTENTIAL.
 FT PROPEP 30 253 BY SIMILARITY.
 FT CHAIN 254 1211 ADAMTS-2.
 FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 409 409 BY SIMILARITY.
 FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 480 560 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 561 617 DISINTEGRIN-LIKE.
 FT DOMAIN 618 722 TSP TYPE-1 1.
 FT SITE 691 693 CYS-RICH.
 FT DOMAIN 723 851 CELL ATTACHMENT SITE (POTENTIAL).
 FT DOMAIN 852 911 SPACER.
 FT DOMAIN 912 974 TSP TYPE-1 2.
 FT DOMAIN 975 1030 TSP TYPE-1 3.
 FT DOMAIN 40 43 TSP TYPE-1 4.
 FT DOMAIN 185 188 POLY-GLU.
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 993 993 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1098 1098 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1145 1145 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1150 1150 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 544 566 HGFCHCIWLTPTDLKRDGSKA -> FRPGVAHACYP5
 FT TILGGGRWIA (IN ISOFORM SPNP1).
 FT VARSPLIC 567 1211 MISSING (IN ISOFORM SPNP1).
 FT SEQUENCE 1211 AA; 134722 MW; BECEE25C23CAD2D CRC64;
 Query Match 24.9%; Score 788.5; DB 1; Length 1211;
 Best Local Similarity 34.9%; Pred. No. 1.1e-52;
 Matches 217; Conservative 79; Mismatches 246; Indels 79; Gaps 27;
 QY 9 LAFAGRTAG---GFEPEREYVVPVIRLDPD-----INGRR-----YYW 42
 DB 39 LAAADPPGGPLGHGAERILAVPVRTDAQRLVSHVVSAAVSAGVRRARRAARVTFPSPP 98
 QY 43 RGPEDSGDGLIFQITAFQEDFVHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCF 102
 DB 99 GGNPEPGSHLFYVTVFGRLHLRPNARLVAPGATMEWQGE--KGTRVEPLLGSL 156
 QY 103 YSGDVN--AEPDSFAAVSLCGGLRGAFGYRGAEYVISPPLPNASAPAAQNSOG-AHLQR 159
 DB 157 YVGDVAGLAERASS-VALSNCGLAGLIRMEEEFFLEKGL--AAQAEQGRVHVYR 213
 QY 160 RGVPGSPSGDPTSCRCVAGSNPAIRLALDYPKPRRAGGSGRRSRKRKRVISIPRY 219
 DB 214 RPTSPPLGPGQA---LDTG---ASLSDLSLS-RALGVLEEHSRRRRRAHADDY 266
 QY 220 -VETLVVADSMVKPHGAD-LEHYLLTLATAARLYRHPISILNPINIVVVKLLRDRDS 277
 DB 267 NIEVLGVDDSVVQFHGKEHVOKYLLTLMNINEYHDESIGAHINVLVRLILSYGKS 326
 QY 278 GPKV-TGNAALTLRNFCAMQKKLNKYSVKHPEYWDATILFTQDLCGATTCDTLGMADVG 336
 DB 327 MSLEIGNPSQSLENVCRWAYLQKQPDTHGDEYHDAIFLTQDF-GPSGMO--GYAPVT 383
 QY 337 TWCDPRKSCSVIEDDGLPSAFTTAHELGHVFNPHDNV--KVCEEVFGKLRANHMMSPTLI 395

Db 384 GMCHPVRSCTLNHEDGFSAFVVAHETGHVIGMEHDGQNRCDG---EVRIGSIMAPLVQ 440
QY 396 QIDRANPWSACSAIITDFDLSHGDCILDDP---SKPISLPEDLPASYSITLSQOCELAF 452
Db 441 AAFHFRHWSRCSQOELSYLHS--YDCLLDDPFAHDWP-ALPQ-LPLGLHSYNEQCRDF 496
QY 453 GVGSKPCPYMY---CTKLWCTKAKGQMVQOTRHPFMDGTSCGBGKCLCKGACV--- 505
Db 497 GLGYMMCTAFRTDPDCKQLWCS-HPDNPFCKTKGKPPDLGTWCAPKCKFCGHCILWTP 555
QY 506 ---ERHNLNKRHVDGSAKWDPTGPGSRCTCGGQVQLARRQCTNPTPANGKYGCEGVVVKY 562
Db 556 DILKR-----DGSWAGSPGSCSRTCTGKVGKTRQCDNPHNPANGRTCSGLAYDF 607
QY 563 RSNLEPCPSSASGKSFREQ 583
Db 608 QLCSDQDCPDLSA--DFREQ 626

RESULT 15
ID AT52_BOVIN STANDARD; PRT; 1205 AA.
AC P79331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-19)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI).
GN ADAMS2 OR NPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=9725960; PubMed=9122202;
RA Collige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,
RA Lapierre C.M.;
RA "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT new member of the superfamily of zinc-metalloproteinases with binding
RT sites for cells and other matrix components.";
RL proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=95348096; PubMed=7622483;
RA Collige A., Beschlin A., Samyn B., Goebels Y., Van Beeumen J.,
RA Nusgens B.V., Lapierre C.M.;
RT "Characterization and partial amino acid sequencing of a 107-kDa
RT procollagen I N-proteinase purified by affinity chromatography on
RT immobilized type XIV collagen.";
RL J. Biol. Chem. 270:16724-16730(1995).
CC -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC COLLAGEN BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC at Ala-1-Gln.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC COLLAGEN TYPE XIV.
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT

CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- DISEASE: DEFECTS IN ADAMS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMS3.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X96389; CAA65253.1; -.
CC MEROPS: M12.301; -.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_Mtpeptdse.
DR Pfam: PF00090; Tsp1; 4.
DR Pfam: PF01421; Reprolysin; 1.
DR SMART: PF01562; Pep_M12B_propep; 1.
DR SMART: SM00209; TSP1; 4.
DR PROSITE: PS50215; ADAM_MEPRO; 1.
DR PROSITE: PS50092; TSP1; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28
FT PROPEP 29 253
FT CHAIN 254 1205
FT METAL 402 402
FT ACT_SITE 403 403
FT METAL 406 406
FT METAL 412 412
FT DOMAIN 474 554
FT DOMAIN 555 611
FT DOMAIN 612 716
FT DOMAIN 717 845
FT DOMAIN 846 905
FT DOMAIN 906 968
FT DOMAIN 969 1024
FT SITE 585 687
FT DOMAIN 31 35
FT DOMAIN 177 180
FT CARBOHYD 104 104
FT CARBOHYD 245 245
FT CARBOHYD 942 942
FT CARBOHYD 943 943
FT CARBOHYD 987 987
FT CARBOHYD 1025 1025
FT CARBOHYD 1092 1092
FT CARBOHYD 1139 1139
FT CARBOHYD 1144 1144
SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;
Query Match 24.5%; Score 775; DB 1; Length 1205;
Best Local Similarity 32.9%; Pred. No. 1.2e-51;
Matches 213; Conservative 82; Mismatches 246; Indels 106; Gaps 26;
QY 1 MLLGLITLAFAGRTAG-----GPEPEVVVPIRLD----- 32
Db 16 LLLLLLPADARLAAAAAAPPFGPGHGAERILAVPVRTDAQGRVSHVVSATAPAG 75

```

QY 33 -----PDINGRRYYWRGPEDSGDQGLIFOITAFQEDFYHLHTPDQAFLAPAFST 81
Db 76 VITRRAAPAIQGLSG-----GSEEDGGRLFYNTVTFGRDLHLRLRNARLVAAGATV 129
QY 82 EHLGVPLQGLTGSGSDLRRCFYSGDV-NAEPDSFAAVSLCGGLRGAFGYRGAEYVISPLP 140
Db 130 EWQGE--SGATRVPLLGTCLYXGDVAGLAESSVALSCDGLAGLIRMEEEFFIEPLE 187
QY 141 NASAPAAQRNSOG-AHLQRRGP-----GGPSGDPISRCGVASGWNPAILRALDPYKP 193
Db 188 KGL--AAKEAQGRVHVYHRTTTRPPPLGGPQALDT---GISADSLDSLRL----- 237
QY 194 RRAEGESRRRRSGRAKRFVSIPIRY-VETLVVADESMVKFHGAD-LEHYLLTLATAAR 251
Db 238 ---GVLEERVNSSRRMRRAADDYNIIEVLGVDSDVQPHGTEHVOKYLLTLMNIVNE 294
QY 252 LYRHPISILNPINIVVVKVLLLRDRSGPKV-TGNAALTLRNFCAWQKKLNKVSOKHPEYW 310
Db 295 IYHDESLGAHINVLVRIILISYKGSMSLIEIGNPSQSLNVCRWAYLQKQKPDTHDEYH 354
QY 311 DTAIFLTRODLGCAHTCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMP 370
Db 355 DRAIFLTRODF-GPSGMQ--GYAPVTGMCHPVRSCTLNHEDGFSFAFVVAHETGHLGME 411
QY 371 HDNV-KVCEEVEGKLRANHMMSPTLIQIDRANPWSACSAAITDFDLSGHGDCLLDQP-- 427
Db 412 HDGQGNRCGD---EVRLGSIAPLVQAAFRHFWNSRCSQOELSRYLHS--YDCLRDDEFT 466
QY 428 -SKPISLPEDLPASYSYLSQOCELAFGVGSKPCPYMQY---CTKLWCTGKAKGMVCQTR 483
Db 467 HDWP-ALPQ-LPGLHYSNEQCRFDGLGYMMCTAFRTFPCQLWCS-HPDNEYFCKTK 523
QY 484 HFPWADGTSCGEGKCLKACV-----ERHNLNKHVRVDSWAKWDPYGPCSRTC GG 536
Db 524 KGPPLDGTWCAFGKCFKGHCILWLPDILKR-----DGNWAWSPFGSCSRTC GT 575
QY 537 QLAREQCTNPTPANGKYCEGVVRYKRSCLNEPCPSSASGKSFREEQ 583
Db 576 KPTRQCDNPHANGRTCSGLAYDFQLCNSQDCPDALA--DFREEQ 620

```

Search completed: April 29, 2003, 17:14:04
Job time : 11.1802 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 35.5509 Seconds
(without alignments)
3378.970 Million cell updates/sec

Title: US-10-009-332-1-copy_1_583
Perfect score: 3163
Sequence: 1 MLLGLITLAFAGRTAGGFE.....SCNLEPCPSASGKSFREQ 583

Scoring table: . BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-rvirus:*
- 16: sp-bacteriap:*
- 17: sp-archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3155	99.7	950	4 Q8TE58	Q8TE58 homo sapien
2	941.5	29.8	2165	5 Q19791	Q19791 caenorhabdi
3	865.5	27.4	1054	5 Q9W493	Q9W493 drosophila
4	803.5	25.4	1095	4 Q8TE56	Q8TE56 homo sapien
5	796	25.2	269	6 Q9GL54	Q9GL54 oryctolagus
6	793	25.1	1081	4 Q8TE60	Q8TE60 homo sapien
7	790	25.0	1688	5 Q8SX80	Q8SX80 drosophila
8	762	24.1	1072	4 Q8TE57	Q8TE57 homo sapien
9	757.5	23.9	1229	5 Q9VF61	Q9VF61 drosophila
10	753.5	23.8	1207	4 Q8TE59	Q8TE59 homo sapien
11	715.5	22.6	1223	4 Q8WX58	Q8WX58 homo sapien
12	715.5	22.6	1223	4 Q8TE55	Q8TE55 homo sapien
13	714	22.6	1159	4 Q8TE58	Q8TE58 homo sapien
14	568.5	18.0	1427	4 Q96L37	Q96L37 homo sapien
15	558	17.6	192	6 Q95N24	Q95N24 equus cabal
16	518	16.4	187	6 Q95N23	Q95N23 equus cabal

17	444	14.0	1091	5 Q9WL26	Q9WL26 drosophila
18	442.5	14.0	1062	5 Q19204	Q19204 caenorhabdi
19	433	13.7	1444	5 Q17591	Q17591 caenorhabdi
20	393	12.4	872	5 Q22580	Q22580 caenorhabdi
21	357.5	11.3	790	5 Q8T458	Q8T458 drosophila
22	335.5	10.6	899	13 Q8UVF1	Q8UVF1 coturnix co
23	332.5	10.5	479	13 Q9PWJ0	Q9PWJ0 agkistrodon
24	331	10.5	922	13 Q8UVF2	Q8UVF2 coturnix co
25	326	10.3	466	13 Q9TAX7	Q9TAX7 agkistrodon
26	322	10.2	477	13 Q98SP2	Q98SP2 bothrops ja
27	319.5	10.1	479	13 Q9PW78	Q9PW78 agkistrodon
28	314.5	9.9	609	13 Q9W6M5	Q9W6M5 agkistrodon
29	313.5	9.9	610	13 Q93523	Q93523 bothrops ja
30	313.5	9.9	610	13 Q8QG88	Q8QG88 bothrops in
31	310	9.8	476	13 Q9YI19	Q9YI19 agkistrodon
32	310	9.8	482	13 Q9PVK9	Q9PVK9 agkistrodon
33	309.5	9.8	610	13 Q9YI20	Q9YI20 agkistrodon
34	309	9.8	617	13 Q90499	Q90499 echis pyram
35	307	9.7	411	13 Q92031	Q92031 agkistrodon
36	307	9.7	466	13 Q9TAX6	Q9TAX6 agkistrodon
37	304	9.6	407	13 Q92032	Q92032 agkistrodon
38	303	9.6	108	6 Q9GLK6	Q9GLK6 oryctolagus
39	303	9.6	483	13 Q9IAB0	Q9IAB0 agkistrodon
40	302.5	9.6	609	13 Q90282	Q90282 crotalus at
41	300	9.5	620	13 Q42138	Q42138 agkistrodon
42	299.5	9.5	400	13 Q9IAY0	Q9IAY0 agkistrodon
43	298.5	9.4	825	11 Q8R3D3	Q8R3D3 mus musculu
44	298	9.4	117	11 Q8VHK4	Q8VHK4 mus musculu
45	297	9.4	549	13 Q90500	Q90500 echis pyram

ALIGNMENTS

RESULT 1

Q8TE58	ID	Q8TE58	PRELIMINARY;	PRT;	950 AA.
AC	Q8TE58;				
DT	01-JUN-2002 (TrEMBLrel. 21, Created)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)				
DE	Metalloprotease disintegrin 15 with thrombospondin domains.				
GN	ADAMTS15.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9605;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-21856482; PubMed-11867212;				
RA	Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,				
RA	Lopez-Otin C.;				
RT	"Cloning, expression analysis, and structural characterization of				
RT	seven novel human ADAMTSs, a family of metalloproteinases with				
RT	disintegrin and thrombospondin-1 domains.";				
RL	Gene 283:49-62(2002).				
DR	EMBL; AJ315733; CAC86014.1; -.				
KW	Integrin; Protease.				
SQ	SEQUENCE 950 AA; 103286 MW; 5DFBE18285CCCC3B CRC64;				
	Query Match 99.7%; Score 3155; DB 4; Length 950;				
	Best Local Similarity 99.8%; Pred. No. 1.3e-253;				
	Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	1	MLLGLITLAFAGRTAGGFEPEPEVVVPIRLDPDINGRRYRWGPDSDGQGLIFQITAF	60		
Db	1	MLLGLITLAFAGRTAGGFEPEPEVVVPIRLDPDINGRRYRWGPDSDGQGLIFQITAF	60		
QY	61	QEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGGSSDLRCFYSGDVNAEPDSFAVSLC	120		
Db	61	QEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGGSSDLRCFYSGDVNAEPDSFAVSLC	120		
QY	121	GGLRGAFGRGAEYVISPLPNASAPAAQNSQGAHLQRRVPGGPGDPTSRGCVASGW	180		

```
Db 121 GGLRGAFYGAAYVISPFPNAPSAQAORNSOGAHLQRRGVPGPSGDPTRSCGVASGW 180
QY 181 NPAILRALDPYKPRRAGFESRRSRRGRKRFVSPRIVETLVVADESVMKVFHGDLEH 240
Db 181 NPAILRALDPYKPRRAGFESRRSRRGRKRFVSPRIVETLVVADESVMKVFHGDLEH 240
QY 241 YLTLTATARLYRHSILNPINIVVVKLLDRDSDGPKVTGNALTLRNFCAWQKLN 300
Db 241 YLTLTATARLYRHSILNPINIVVVKLLDRDSDGPKVTGNALTLRNFCAWQKLN 300
QY 301 KVSQKHEPYWDTAILTRQDLGCATTCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTA 360
Db 301 KVSQKHEPYWDTAILTRQDLGCATTCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTA 360
QY 361 HELGHVFNPHDNVNVKVEEYFVKLRANHMSPTLIOIDRANPWSACSAAIITDFDSHG 420
Db 361 HELGHVFNPHDNVNVKVEEYFVKLRANHMSPTLIOIDRANPWSACSAAIITDFDSHG 420
QY 421 DCLLDQSPKISLPEDLPASVYLSOCELAFGVGSKPCPYMOYCTKLMCTGKAKQWVC 480
Db 421 DCLLDQSPKISLPEDLPASVYLSOCELAFGVGSKPCPYMOYCTKLMCTGKAKQWVC 480
QY 481 QTRHFPWADTSCGEGKCLKLGACVVRHNKHRVDGWSNAKWDYPGCSRTCGGGVQLAR 540
Db 481 QTRHFPWADTSCGEGKCLKLGACVVRHNKHRVDGWSNAKWDYPGCSRTCGGGVQLAR 540
QY 541 RQCTNPTPANGKYCEGVVRYKNSCNLEPCPSSASGKSFREQ 583
Db 541 RQCTNPTPANGKYCEGVVRYKNSCNLEPCPSSASGKSFREQ 583

RESULT 2
Q19791
ID Q19791 PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RA Gjadasty S.;
RP SEQUENCE FROM N.A.
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Snaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wooldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1; -.
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1; -.
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSSP; P15167; IDTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPL.
```

```
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp.1; 14.
DR SMART; SM00209; TSPL; 18.
DR PROSITE; PSS0215; ADAM_MEPRO; 1.
DR PROSITE; PSS0092; TSPL; 6.
DR PROSITE; PSS0142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;

Query Match 29.88; Score 941.5; DB 5; Length 2165;
Best Local Similarity 38.78; Pred. No. 5.3e-69;
Matches 203; Conservative 75; Mismatches 181; Indels 65; Gaps 14;

QY 101 CFYSGDVNAEPDSPAFAVSLC---GGLRGAFYGAAYVISPFPNAPSA---PAAQRNSQGA 154
Db 169 CIYRAHYKG-VHQHSIVNLCDSDEGLYGLMALPSGIHTVEPIISGNGTEHDGASRRHQ-- 225
QY 155 HLQARRGVGPGSDPTSRGCV-----ASGNPAILRALDPYKPRAGGESRSR 204
Db 226 HLVRKFDPMHFKSFDHLNSTSVNETTVAWQDQWEDVIER-----KARSR 272
QY 205 RRSRAKRFVSIPIRVETLVVADESVMKVFHGDLEHLLTLLATAARLYRHPISILNPINI 264
Db 273 RAAN-----SWDHYVEVLVVDATKMYEYHGRSLEDIVLTSTVASIYRHSQSLASINV 326
QY 265 VVYKVLRLDRDSDGPKVTGNALTLRNFCAWQKLNKVKSDKHPYWDTAILTRQDLGCA 324
Db 327 VVVKLIVLKTENAGPRITQNAQQTLDQFCRWQYINPDQSSVQVHDDVAIILLTRKDIERS 386
QY 325 T-TCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKVEEV--- 380
Db 387 QGKCDTLGLAELGTMDCMQKSCAIIEDNGLSAAFTIAHELGHVFSIPHDDERKSTYMPV 446
QY 381 -----FCKLRAN---HMSPTLIOIDRANPWSACSAAIITDFLDSHG--DCLLD 425
Db 447 NKVKFQSTKEDKTOFQNNFHIMAPTLEYNTHPWSWSPCSAGMLERLENNRGQTQCLFD 506
QY 426 QPSKISLPE---DLPGASYLSOCELAFGVGSKPCPYMOYCTKLMCTGKAKQWVCQ 481
Db 507 QPVERRYVEDVVRDEPGKKYDAHQCKFVFGPASELCPYMPTCRLMLCATFYQSWMGCR 566
QY 482 TRHFPWADTSCGEGK---LCLKLGACVVRHNKHRVDGWSNAKWDYPGCSRTCGGGVOLA 539
Db 567 TQHNFPWADTSCDESRSMFCHGACVRLAPESLTKIDQWGDWRSWGSBSCSRTCGGGVQKG 626
QY 540 RQCTNPTPANGKYCEGVVRYKNSCNLEPCPSSASGKSFREQ 583
Db 627 LRDCSDPKPRNGKYCVGQGRYRSCNTQECPWDT--QPYREVQ 668

RESULT 3
Q9W493
ID Q9W493 PRELIMINARY; PRT; 1054 AA.
AC Q9W493;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CG4096 protein.
GN CG4096.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
```

QY	202	RSRRSRGRKRFVSIPIRYVETLVVADBSWVKFHGADLEHYLLTLLATAARLYRHPISILNP	261
Db	309	PDHAEIPRRRSISSPRHVETLIVADATMGAFH-RDLNGYLLTIMNVSALYKDPDSIGNS	367
QY	262	INIVVKVLLLRDSDGPK--VTGNAALTLRNFCAMQKLNKYSKDKHPYWDTAIFLTRQ	319
Db	368	IEIVVVKRIQDDEESQLNLNTQNAQKNLDRFCWQHKLNKSEKDPHHHDVAAILTRK	427
QY	320	DLGATTCDTLGMADVGTMDPKRRCSVIEDDGLPSAFTTAHELGHVFNPNPHDNVAV-CE	378
Db	428	NIC-ANNCMTLGLANVGCMCKPKQSCSVNEDNGIMLSHTITHELGHNFMGHDTAKIGCH	486
QY	379	EVFGKLRANHMSPTL-----IQIDRANPWSACSAAIITDFLDGSHGDCLLDQPSKPI---	431
Db	487	PRVGPI--VHIMTPTFGADTLQV-----CWGNCRSKRYITHELDQGLGEC-LDDPPTLDEY	539
QY	432	SLPEDLPASVTLSSQCELAFG-----VGSKPCPYMOCYCTKLWCTGKAKGOMVCOIRHF	485
Db	540	NYTGELPGMYNARGCQRLQFNLTITDSEVGACSAHP-EFCSTLWC--KYNGE--CVTHMR	594
QY	486	PWADGTCGEGKLCGLKACAYVERHNLNKHRYDGVSAKWDPYGPCSRTCGGQVLARRQCTN	545
Db	595	PTAPGTLGRNWKQCKVCYRREEL--AAVNGMGDWSEWSECSRSCGGGVSTQORECDN	652
QY	546	PTPANGKICYCEGVRYKYRNSCNLEPCPSSASGKSFREQ	583
Db	653	PVPANGVFCIGERKRYKICKRKPCC--AEEPSFRAQQ	688
RESULT 4			
Q8TE56	ID	Q8TE56 PRELIMINARY; PRT; 1095 AA.	
AC	Q8TE56	AC Q8TE56	
DT	01-JUN-2002	(TEMBLrel. 21, Created)	
DT	01-JUN-2002	(TEMBLrel. 21, Last sequence update)	
DT	01-JUN-2002	(TEMBLrel. 21, Last annotation update)	
DE	DE	Metalloprotease disintegrin 17, with thrombospondin domains.	
GN	ADAMTS17.		
OS	Os	Homo sapiens (Human).	
OC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=21856482; PubMed=11867212;		
RX	Cal S., Obya A.J., Llamazares M., Garabaya C., Quesada V.,		
RT	Lopez-Otin C.;		
RT	"Cloning, expression analysis, and structural characterization of		
RT	seven novel human ADAMTS, a family of metalloproteinases with		
RT	disintegrin and thrombospondin-1 domains.";		
RL	Gene 283:49-62(2002).		
RL	EMBL; AJ315735; CAC86016.1; -.		
KW	Integrin; Protease.		
DW			
SQ	SEQUENCE 1095 AA; 121099 MW; A5C6540484754D5D CRC64;		
Query Match 25.4%; Score 803.5; DB 4; Length 1095;			
Best Local Similarity 33.5%; Pred. No. 6.2e-58;			
Matches 212; Conservative 93; Mismatches 243; Indels 85; Gaps			
QY	2	LLILGLTAF-----AGRTAGGFGPERVVVPIRLDPI-----NGRRYYVRGPE	46
Db	10	LVLPLVLLLVGLDPCGTAVGDAADVEVLVPRVPRDDVHLPLPAAPGPRRRRPTTP	69
QY	47	DS-----GDGLIPQITAFQEDYELHITPDQAFLAPAFSTEHLGVLPQGTGSSDLRRC	101
Db	70	AAPRARPERALLHLHPAFGRDLYLQLRRDLRLSLRSGFEVEEAGAARR--RGRPAEL--C	125
QY	102	FYSCDVNAEPDSPAAYSLCGLRGATG--YRGAAYV-ISP LNPASAPAAQRNSQGAHILQ	158
Db	126	FYSGRVLGHPSGLVSUSACGAAGGLVGLIQIQGEQVLIQPLNNSQGFPSGRE-----HLIR	181
QY	159	RR-GVPGGPGSD---PTSRGCVASGWNPAIILRALDPLVKPRACFGE-SRSRRRSRAKRF	213

Db	182	RKWSLTSPSAEAAQRPEQLCKVLT-----EKKKPTWGRPSRDWRERRNAIRL	228
Qy	214	VSPRYVETLVVADESAAVKGHGD-LEHYLLTLTAAARLYRHPSILNPINIVVKVLLL	272
Db	229	TS-EHTVETLVVADADWQYHGAEAAQRFILTVNWMVYNMFQHSIGIKINIQVTKLVLL	287
Qy	273	RDRDGGPKVTGNAALTLRNFCAMQKK-----LNKV--SDKHPEYWDTAILFTRODL	321
Db	288	RQRPAKLSIGHGERSLESFCHWQNEEYGGARYLGNQVPGDKDDPLVDAAEVTVRTDF	347
Qy	322	C--GATTCDFLGMADVGTMCDPKRSQSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEE	379
Db	348	CVHKDEPCDVTGVIAYLGGVGSARKKCVLAEDNGLNLAFITAEHLGHNLGMNHD-----DD	402
Qy	380	VFGKLRANHMSPTLIQIDRANP---WSACSAAILTDFLSDHGHGDCLL--DQSPKPS	432
Db	403	HSCAGRSHMSGBWKA--GRNPSDLSSWSSCRDDELFLKSVKSTCLLYTDPQRSQHTVR	460
Qy	433	LPEDLPASVTLTQQCELAFGVSKPCPYMQY--CTKLWCTGKAKGOMVQOTRHFPPWADG	490
Db	461	LPHKLPGMHSYANECQQLFGMNATFCRNHEHLMCAGLWCL--VEGDTSCKTGLDPPLDG	518
Qy	491	TSCGEGKCLKLGACVERHNLNKHRVDSWAKWDYPGCSRTCGGGVOLARQCTNPTPAN	550
Db	519	TEGADKWCRAEGCVSKTPIPEH--VDGDWSPWGAWSMCSRTCCTGARFRQKCDNPPPGP	577
Qy	551	GGKYCEGVRYKRYSCNLEPCPSSASGSKFRREQ	583
Db	578	GGTHCPGASVEHAVCENLPCPKGL--PSFRDQQ	608
RESULT 5			
Q9GL54			
ID	Q9GL54	PRELIMINARY;	PRT; 269 AA.
AC	Q9GL54;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Aggrecanase-2 (Fragment).		
GN	ADAMTS-11.		
OS	Oryctolagus cuniculus (Rabbit).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.		
ON	NCBI_TaxID=9986;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RA	Goat D.L., Goat M.E.;		
RT	"Molecular cloning of lapine aggrecanase-2 (ADAMTS-11) from articular chondrocytes."		
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF317415; AAC33062.1; -		
DR	HSSP; Q9PW35; IBDU.		
DR	MEROPS; M12.225; -		
DR	InterPro; IPR001590; Reprolysin.		
DR	InterPro; IPR000884; TSPI.		
DR	InterPro; IPR000130; Zn_MTpeptidse.		
DR	Pfam; PF01421; Reprolysin; 1.		
DR	Pfam; PF00090; tsp_1; 1.		
DR	SMART; SM00209; TSPI; 1.		
DR	DR PROSITE; PS00215; ADAM_MPEPRO; 1.		
DR	DR PROSITE; PS00092; TSPI; 1.		
DR	DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.		
FT	NON_TER		
FT	NON_TER		
FT	SEQUENCE	269 AA; 29193 MW; 97AICA80B33452FA CRC64;	
Query Match			
Best Local Similarity 53.8%; Score 796; DB 6; Length 269;			
Matches 140; Conservative 37; Mismatches 73; Indels 10; Gaps 3;			
Qy	329	TLGMADVGTMCDPKRSQSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEEVFGKLRANH	388

Db	1	TIGMADVGTCTSPERSCAVIEDDGLHAAFTVAHEIGHLLGSLSHDDSKFCEENFGSTEDKR	60
Qy	389	MMSPFLIIDRANPWSACSAAITDFLDGSHGDCLLDQPSKPISLPEDLPAGASYTLUSQQC	448
		: : : : : : : : : : : : : : : : : :	
Db	61	LMSSILTSIDASKPWCKTSAITTEFLDDHGNCCLLDVPRKQTILGPPEELPGOTYDATQC	120
Qy	449	ELAFGVSGKCPPTMYQYCTKLWCTGAKAQMWCVQTRHPFNADGTSCEGKLCCLKGACVER-	507
		: : : : : : : : : : : : : : : : :	
Db	121	NLTFGPEYTPCGMDVCARLWCRAVVRQGQWCVTKKLPAVEGPPCGKGRICLQGGKVDKT	180
Qy	508	----HNLNKHRRVGDSNAKWDPYGPCSRTCGGVLARRQCTNPTPANGKYCEGVVKYR	563
		:::: : : : : : : : : : : : : : : : : :	
Db	181	KKKYITSSSH---GNMGSWGPGSCRS CGGGVQFAFYRHCHNNPAPRNRYCTGKRAYIR	237
Qy	564	SCNLPCPSSASGKSFRREQ	583
		: : : : :	
Db	238	SCSVTPCP--ANGKSFREQ	255
RESULT 6			
ID QTE60		PRELIMINARY; PRT; 1081 AA.	
AC QTE60;			
DT 01-JUN-2002 (TREMBLrel. 21, Created)			
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE ADAMTS18 protein.			
OS ADAMTS18.			
GN Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=21856482; PubMed=11867212;			
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,			
RA Lopez-Otin C.			
RT "Cloning, expression analysis, and structural characterization of			
RT seven novel human ADAMTSs, a family of metalloproteinases with			
RT disintegrin and thrombospondin-1 domains."			
RL Gene 283.49-62(2002).			
DR ENBL; AJ311903; CAC83612.1; -			
SQ SEQUENCE 1081 AA; 119655 MW; 0438BF645676461E CRC64;			
Query Match	25.1%;	Score 793;	DB 4; Length 1081;
Best Local Similarity	32.8%;	Pred. No. 4.6e-57;	
Matches 211; Conservative	85;	Mismatches 229;	Indels 118; Gaps
Qy	16	AGGFEPEREVVVPIRLDP-----DI--NGRRYYWRGPDSDGGQLIFQITAFQEDFYL	66
	:	:	:
Db	54	ASGLNDXVFVTPPEVDSAGSYISHDLTHNGRK---KRSQAQNARSLSHYRFSAFGEHLH	110.
Qy	67	HLPDAQFLAPAFSTEHLGYPLQGLTGSS-----DLRRCFYSGDVNAPDSPFAAVSLCGG	122
	:	:	:
Db	111	ELKFS-----AIIUSHFIVOVLGKGASETQKPEVQCFQGIIRDSSSSVAVSTCAG	164
Qy	123	LRGAGFYRGAEVVISPLP-----NASAPAQRNSOGAHLLOR-----RGYPG-	164
	:	:	:
Db	165	LSGLIRTKNEFLISPLPOLLAEHHNYSSPAGHH---PHVLYKRTAEKIQRVGVPGS	220
Qy	165	-----GPSGDP---TSR-----CGVASGWNPAILRALDPKP-----	193
	:	:	:
Db	221	GRNYPGVSPSHIPASQSRETEYHHRLLQKHQFCGRKKYAP-----KPTEDTYL	271
Qy	194	RRAFGES-RSRRESGRAKREVSPRVETLVADESMVAFHG-ADLEHVLLLTLLATAAR	251
	:	: : : : : : : : : : : : : : : : :	
Db	272	RFDYIGSGSRPRRSAGKSQGLN----VETLVADKKMKVEKHGKNVTIILVVMKVSG-	326
Qy	252	LYRPSILNPINIVVKKVLLDRDSDGPKYTGNAALTRLNFCAWQKKL-NKVSOKHPEYW	310
	:	: : : : : : : : : : : : : : : : :	
Db	327	LFKDGTIGSDINVVVSVLILLEQEPGGLLIINHDAQSLNSEFCQWSALICKNGKRH---	382
Qy	311	DTAIFTLRQDLCG--ATTCDTLGMADVGTMC DPRCSVIEDDGLPSAFTTAHELGHVN	368

Db	556	ICHMLGAEELGTVC\$-SSSC\$IVQDTGLPTAFTMAHELGHILNNHHDDDDKCM\$PVYTRQN	614
Qy	386	AN---HMSPTLIQIDRANPWSACSAALIITD\$D\$HGDC\$LLDQ\$PK\$PI\$P---EDLP\$	439
Db	615	NNKVLH\$MSVNGTHMHPW\$SK\$SRHFVS\$E\$K\$TK\$CL--ETS\$G\$A\$H\$P\$Y\$T\$ER\$LP\$	672
Qy	440	ASYT\$LSOOC\$ELAF\$G\$V\$K\$P\$Y\$M\$Y\$C\$TK\$LC---TG\$K\$AQ\$M\$V\$C\$Q\$TR\$H\$F\$W\$AD\$T\$G\$C\$-E	495
Db	673	EI\$Y\$LD\$AQ\$CL\$F\$N\$D\$F\$C\$Y\$P\$T\$D\$E\$C\$K\$RL\$M\$C\$N\$RT\$S\$G\$N\$E\$Q---C\$A\$S\$N\$LP\$W\$AD\$T\$P\$C\$G\$S\$	730
Qy	496	GK\$CL\$K\$AC\$Y\$E\$R\$H\$N\$K\$H---R\$V\$D\$G\$W\$AK\$W\$D\$P\$G\$C\$S\$R\$T\$C\$G\$G\$V\$OL\$AR\$Q\$C\$N\$P\$T\$P\$ANG	551
Db	731	GH\$W\$C\$Q\$K\$C\$V\$S---N\$K\$H\$G\$Y\$R\$Q\$V\$N\$G\$W\$G\$P\$W\$T\$P\$T\$P\$C\$SL\$T\$C\$G\$G\$V\$Q\$E\$S\$R\$E\$C\$N\$Q\$P\$V\$P\$ENG	786
Qy	552	KY\$C\$E\$G\$V\$V\$K\$Y\$R\$S\$C\$N\$LE\$P\$C\$P\$S\$S\$A\$G\$K\$S\$F\$R\$E\$Q	583
Db	787	KY\$T\$G\$R\$K\$K\$Y\$R\$S\$C\$N\$T\$H\$Q\$C\$P\$G\$M\$D\$P---R\$E\$Q	816

RESULT 8

ID	Q8TE57	PRELIMINARY; .	PRT; 1072 AA.
AC	Q8TE57;		
DC	01-JUN-2002 (TReMBLrel. 21, Created)		
DT	01-JUN-2002 (TReMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)		
DE	ADAMPTSL6		
GN	metalloprotease disintegrin 16 with thrombospondin type I motif.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
ON	NCBI_TaxID=9606;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21856482; Pubmed=11867212;		
CA	Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,		
RA	Lopez-Otin C.		
RT	"cloning, expression analysis, and structural characterization of		
RT	seven novel human ADAMTSs, a family of metalloproteinases with		
RT	disintegrin and thrombospondin-1 domains.";		
RL	Gene 283:49-62(2002).		
RL	DB EMBL; AJ315734; CAC86015.1; .		
RT	Integrin; Protease.		
SK	SEQUENCE 1072 AA; 119491 MW; 081EEFF78F47D061 CRC64;		

Best Local Similarity 34.0%; Pred. No. 1.7e-54;
Matches 200; Conservative 74; Mismatches 227; Indels 88; Gaps 21;
QV 51 OGLFIOTAFQEDFVLIHTPDAAOFIADAFSTFHTGVPIQIGTAGCS-----DDEGV 103

Db	95	ESLHLRLKGRPHDFHMDLRTSSSLVAPGFIQVTL-----GKTCTKSVQTLPPEDF--CFY 147
QY	104	SGDVNRPDSFAVSLCGGLRGURGAFGRGAEVYISPLPNASAPAAQNSOG---AHLQRR 160
Db	148	QGLSRHRNSVSVALSTCGQLSGMRITEEADYFLRPLFSLHSLKGLRAAGCGSSPHVLYR 207
QY	161	G-----VPGGPGSDPTSR-----GQVASGWNPA-----ILRA 187
Db	208	STEPHAPGASEVLVTSRWELAHQIHLSSDLRLGLPKQHFGRKKYWPQPKEDFLTL 267
QY	188	LDYKPRAGFGSGSRRRSGRAKRFVSIPIRYVETLWVADESMVKFHG-ADLEHYLLTL 246
Db	268	PDEYK---SCLRHKRSLLRSHNEEL-----NVETLVVDKKMQMNGHENITVLTLL 319
QY	247	ATAALRYRHSILNPINIVVKKVLLLRDRDSGPKVTGNAALTLRNCQAKKL-NKVSDK 305
Db	320	NWYSALFKGTIGCNINIAVLGILLEDSQGLVLSHHADHTLUSFCQWQSGMLMGDKGR 379
QY	306	HPEYWDTAILFTRODLGG--ATTCDTLGMADVGTMDCKPRKSVIEDDGLPSAFTTAHPL 363
Db	380	H----DHAILLTGLDICSWKNEPCDTLGFAPISGMSKYRSCTINEDTGLGLAFTIAHES 435

QY 364 GHVFNNPHDNV-KVCEVEFGKLRANHMSPPTLIQIDRANPNWSACSAALITDPLDSGHGDC 422
 DB 436 GHNFMIHDEGNNCKKSG-----NIMSPITLAGNNGVFSNPSRQYRLHFLSTAQAIC 490
 QY 423 LLDQPSKPI---SLPEDLPASYYTISOQCELAFAFGVSGPCPY---MQVCTKLWC--TGCA 474
 DB 491 LADQP-KPVKEYKYPEKLPGELYDANTOCKWQFGEKAKLCLMDLDFKDKICKALWCHRIGK 549
 QY 475 KGQVCOCTHFWADGTCGEGKCLKAGACVERHNLNHRVDGSKWAKWDPYGPCSRTCGG 534
 DB 550 -----CETRFMPAAGETICGHDMWRCGGCVKYGDEGPKPHGHSDWSSNSPSCSRTCGG 604
 QY 535 GVQLARQCTNPTNPPANGKYCEGVRYKYSNCNLEPCPSASGKSFPREQ 583
 DB 605 GVSHRSLCTNPKPSHGKFCBEGSTRTKLCNSQKCPRDS--VDFRAA 651

RESULT 9
 Q9VF61

ID Q9VF61 PRELIMINARY; PRT; 1229 AA.
 AC Q9VF61;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
 DE CG6107 protein.
 GN CG6107.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberg C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).

DR EMBL: AEO03709; AAF55199.1; --
 DR FlyBase; FBgn0038340; CG6107.
 DR InterPro; IPR002870; Rep_M12B_propep.
 DR InterPro; IPR001590; Repolysin.
 DR InterPro; IPR000884; TSPI.
 DR InterPro; IPR000130; Zn_MTpeptidse.
 DR Pfam; PF01562; Rep_M12B_propep; 1.
 DR Pfam; PF01421; Repolysin; 1.
 DR Pfam; PF00090; tsp_1; 2.
 DR SMART; SM00209; TSPI; 3.
 DR PROSITE; PS0215; ADAM_MEPRO; 2.
 DR PROSITE; PS0092; TSPI; 2.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR Hydrolase; Metalloprotease; Zinc.
 KW SEQUENCE 1229 AA; 138136 MW; C1293180AB19B700 CRC64;
 Query Match 23.9%; Score 757.5; DB 5; Length 1229;
 Best Local Similarity 33.3%; Pred. No. 4.9e-54;
 Matches 191; Conservative 91; Mismatches 184; Indels 107; Gaps 27;
 QY 55 FOITAFQEDFYHLHTPDQAFLAPAFSTEHLGVPLOG-----LTGGSSDLRR--CFYSGD 106
 DB 195 YNLNVFGQLHLVLRDASFY-HNHSMTIRILKEGHEHPGEPETEAEAEQRHLCCFYSGY 253
 QY 107 VNAEPDSFAAAYSLCGGLRGAFYGAAYVISPLPNASAP-----AAQNSOGAHLQ 158
 DB 254 VEDDPHSMVSVSLCGMTGIYKTSFGALLIQPVNRTSSDEVLRHVRKSKORAR--HAYS 311
 QY 159 RRGVGGSGDPTSRGCVASGNPAIRALDPYKPRAGFGE-----SRSRRRSGRAKRFV 214
 DB 312 KPEL-----GLDFMSKLEQVQEEQKSKRLNKKRHYA 347
 QY 215 SIPRVV---ETLVVADESVMKFGHADLEHLLTLTALTAARLYRHPSTLNPINIVVVKLL 271
 DB 348 DVDNOVYTLVLIADVNSMKQFHGEDLQPYILILMSIVSSIFADASGNSTRILLRLIS 407
 QY 272 LRDRDSGPKV---TGNAALTLRNFCAWOKLKNVSKDKHPEYWDYTAFLTRODLGAT--- 325
 DB 408 L-----PNINDOTHSSNMLKHCQF--INQSYER---DTAMLTREPIGCVPGK 454
 QY 326 TCDTIGMADVGMPCKPSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLR 385
 DB 455 ICHMLGLAELGTVCV-SSSCSIVQTLGTFTTAHELGHILNMNHHDDDDCKMPYVTQN 513
 QY 386 AN---HMSPTL-IOIDRANPWSACSAALITDPLDSGHGDCGLDQPSKPISLP---EDLP 438
 DB 514 NNKVLHSMSSVMGIHM---HPMSWSKK---TD-----KSCL--ETSVGAHIPYGTBLP 559
 QY 439 GASVYTLISOCELAFAFGVSGKPCPYMYQYCTKLWC---TGKAGOMVQCTRHFFWADGTCG- 494
 DB 560 GEIYSLDAQCQLSFGNDFCYCPTDECKRLMCNRTSGNSNQ--CASNLPWADGTPCGS 617
 QY 495 EGKCLKLGACVERHNLNKH---RVDSWAKWDYGPCSRTCGGQVQLARQCTNPTAN 550
 DB 618 SGHNCORGCVS-----NRHGYGRQVNGMGWPTPTTSCSLTCGGGQVQESRECNCPVEN 673
 QY 551 GGYKCYGVGVKRVYRSCNLEPCPSASGKSFPREQ 583
 DB 674 GGYCYGTSRKKYRSCNTHOCPPGSMDDP--REQQ 704
 RESULT 10
 Q8TE59 PRELIMINARY; PRT; 1207 AA.
 ID Q8TE59
 AC Q8TE59;
 DT 01-JUN-2002 (TREMELrel. 21, Created)
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
 DE ADAMTS-19.
 GN ADAMTS19.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

A disintegrin-like and metalloprotease with thrombospondin type 1 motif 14 precursor.

DE DE 1
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=21638061; PubMed=11779638;
RX Bolz H., Ramirez A., von Brederlow B., Kubisch C.;
RT "Characterization of ADAMTS14, a novel member of the ADAMTS
RT metalloproteases family";
RL Biochim.Biophys. Acta 1522:221-225(2001).
RL EMBL; AF358666; AAL40229.1; .
DR InterPro: IPR002870; Pep.M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSPI.
DR Pfam; PF01562; Pep.M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 4.
DR SMART; SM00209; TSPI; 4.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPI; 1.
FW SIGNAL; Integrin; Protease; Metalloprotease.
KW SIGNAL 1 22 POTENTIAL.
SQ SEQUENCE 1223 AA; 133871 MW; 36394AC4D92F170F CRC64;

Query Match 22.6%; Score 715.5; DB 4; Length 1223;
Best Local Similarity 31.2%; Pred. No. 1.5e-50;
Matches 175; Conservative 86; Mismatches 238; Indels 61; Gaps

QY 45 PEDSGDGLFIQIFAFQEDFVYLHTPDQAFAPAFSTEH-----LGVPLQGLTGSSD 97
DB 98 PCRVGRHSILYENTVYFGKELHLRLPNRLVVGSSVWQEDFRELFRQPLR----- 149
QY 98 LRRCFYSGDWNABPDSFAAVYLCGGURGAFGYGAEEVYISLPNASAPAAQRNSQG-AHL 156
DB 150 -QECVYTGVTGTPGGAVAISNCDGLAGLRTDSTOFFIEPLRGO---QEKASGRTHV 205
QY 157 LORRGVP---GGPSGDGSRSCGVASGWNPAILRALDPYKPRRAGFGESSRRSRRAKR 212
DB 206 VYREAVQQWAEPPDGLDHNH-AFGLGLDNLNLGLVGQDQ-----LGDYERKRRHAKPGS 258
QY 213 FVSPRYVETLVWADESMVKFHGAD-LEHYLLTLTAAARLYRHPHSILNPINIVVVKVLL 271
DB 259 Y-----SIEVLLVYDDSVRFHGKEHYQNVVLTLMNIVDYIYHDESLGVHINIALVRLIM 313
QY 272 LRDRDSGPKV-TGNAALTILNFCAWQKLNKVSDDKHPEYWDYTAILTRQDLGCATTCDTL 330
DB 314 VGYRQSLSLIERGNPSRSLEQVCRWHAHQQRDPSPSHAEHHDHVYVFLTRQDF-----GPS 367
QY 331 GMADVGTMCDPKRSCEVIEDGLPSAFPTTAHELGHVFNMPHD---NVKVCVEEYFGKLRAN 387
DB 368 GYAPVTGMCHPLRSCALNHEDGSSAFVAIETHGVILGMEHDQGGNCADETSIGS---- 423
QY 388 HMMSPTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQPSKPI-SLPEDLPGASTYLSQ 446
DB 424 -VMAPLVQAFAHFRHFWRSCKLELSRYLPS--YDCLDDPDPFAWPQPELPGINTYSMDE 480
QY 447 QCELAFGVSGKPC---PYMQYCTKLWCTGRAKGMQVCTRHFPWADGTCGEGKCLKGA 503
DB 481 QCREFDGGYQTCIAFTTFEPCOLWCS-HPDNPYFCKTKKGPLDGTGECAPGWCFKGH 539
QY 504 CVERHNLNKHRVDSWAKWDPYGPCSRITCGGGVGLARRQCTNPTTPANGYCYGVRVKYR 563
DB 540 CIWKSPEQTYQDGGWSSWTKFGSCRSCTGGVRSRSCNPNPSPAYGGRPCLGPMEFYQ 599
QY 564 SCNLEPCPSSASGKSFREEQ 583
DB 600 VCNSEECPTGY--EDFRAQQ 617

RESULT 12

```
Q8TE55 PRELIMINARY; PRT; 1223 AA.
ID Q8TE55
AC Q8TE55;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Metalloprotease-disintegrin protease.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=FETAL LUNG;
RA MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ345098; CAC87943.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1223 AA; 133925 MW; D585B6593977ED15 CRC64;

Query Match 22.6%; Score 715.5; DB 4; Length 1223;
Best Local Similarity 31.2%; Pred. No. 1.5e-50;
Matches 175; Conservative 86; Mismatches 238; Indels 61; Gaps 18;

QY 45 PEDSGDGLIFQITAEQEDFYHLHTPDQAFLAPAFSTEH-----LGVPLQGLTGSSD 97
DQ 98 PGRVGRHSLYFNVTGKELHLRLPNRLVPGSSVEMQEDFRELFRQLR----- 149
QY 98 LRRCFYSGDVNAEPDFAAVSLCGGLRGAFYGAEGYVISPPLNAPASAPAAQNSQG-AHL 156
DQ 150 -QECVTGGVTGNGPAGVAISNCDGLAGLIRTDSTDFIEPLERGO---QEKEASGRTHV 205
QY 157 LQRRGVP-----GGPSGDPTSCRCVAGSNPAILRALDYPKPRRAGFESRRSRRAK 212
DQ 206 VYRREAVQGEAEPDGDHLNE-AFGLGLDPLNLLGLVGQD-----LGDTERKRRAKPGS 258
QY 213 FVSIPRYVEVLVADSMVKFHGAD-LEHYLLTLATAARLYRHPSILNPINIVVVKLL 271
DQ 259 Y-----SIEVLLVDDSVYRFHGEHVQNYVLTLMNIVDEIYHDSGLGHINIALVRLIM 313
QY 272 LRDRDSGPKV-TGNAALTLRNFCWAKQKLNKVS DKHPEYWDTAIFTRQDLGCGATTCDTL 330
DQ 314 VGYRQSLSLIERGNPSRSLQVCRWAHSQQRDPSHAEHHDHVVFTRQDF-----GPS 367
QY 331 GMADVGTMCDPKRCSCSVIEDDGLPSAFTTAHELGHVFNMPHD---NVKYCEEVFGKLRAN 387
DQ 368 GYAPVTGMCHPLRSCALNHEDGFSFAFVIAHETGHVLMGHDGQNGCADETSLSGS---- 423
QY 388 HMASPTLIQIDRANPWSACSAAITDFLDSCGDCLLDPSKPI-SLPEDLPGASYTLQ 446
DQ 424 -VMAPLVQAARFHRHWSRCSKLESLRPLS--YDCLDDPDPDPANPQPPPELPGINYSME 480
QY 447 QCELAFGVGSKPC---PYMOYCTKLWCTGKAKGMQVQOTRHPFWADGTSCGEGKLC LKA 503
DQ 481 QCRFDGSGYQTCIAFRTEPCQKLMCS-HPDNYPYFCKTKKGPPLDGTGECAPKWCFKGH 539
QY 504 CVERHNLNKHVRDGSWAKWDYPGCSRTCGGGVQLARRQCTNPTPANGKYCEGVVRKYR 563
DQ 540 CIWKSPEQTYGQDGGWSSWTKFGSCRS CGGVRSRSCNNPSPAYGGRPCGLGPMFEYQ 599
QY 564 SCNLEPCPSSASGKSFREEQ 583
DQ 600 VCNSECEPGTY--EDFRAQ 617
```

RESULT 13

```
Q96L37 PRELIMINARY; PRT; 1427 AA.
ID Q96L37
AC Q96L37;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
```

Q8TEY8

```
ID Q8TEY8 PRELIMINARY; PRT; 1159 AA.
AC Q8TEY8;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ADAMTS14.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21839041; PubMed=11741898;
RA Collige A., Vandenberghie I., Thiry M., Lambert C.A., Van Beeumen J.,
RA Li S.W., Prockop D.J., Lapierre C.M., Nussgens B.V.;
RA "Cloning and Characterization of ADAMTS-14, a Novel ADAMTS Displaying
RT High Homology with ADAMTS-2 and ADAMTS-3";
RL J. Biol. Chem. 277:5756-5766(2002).
DR EMBL; AF366351; AAL79814.1; -.
SQ SEQUENCE 1159 AA; 127336 MW; A5B130149BF7FF34 CRC64;

Query Match 22.6%; Score 714; DB 4; Length 1159;
Best Local Similarity 31.4%; Pred. No. 1.9e-50;
Matches 176; Conservative 87; Mismatches 239; Indels 58; Gaps 19;

QY 45 PEDSGDGLIFQITAEQEDFYHLHTPDQAFLAPAFSTEH-----LGVPLQGLTGSSD 97
DQ 31 PGRVGRHSLYFNVTGKELHLRLPNRLVPGSSVEMQEDFRELFRQLR----- 82
QY 98 LRRCFYSGDVNAEPDFAAVSLCGGLRGAFYGAEGYVISPPLNAPASAPAAQNSQG-AHL 156
DQ 83 -QECVTGGVTGMPGAAVAISNCDGLAGLIRTDSTDFIEPLERGO---QEKEASGRTHV 138
QY 157 LQRRGVP-----GGPSGDPTSCRCVAGSNPAILRALDYPKPRRAGFESRRSRRAK 212
DQ 139 VYRREAVQGEAEPDGDHLNE-AFGLGLDPLNLLGLVGQD-----LGDTERKRRAKPGS 191
QY 213 FVSIPRYVEVLVADSMVKFHGAD-LEHYLLTLATAARLYRHPSILNPINIVVVKLL 271
DQ 192 Y-----SIEVLLVDDSVYRFHGEHVQNYVLTLMNIVDEIYHDSGLGHINIALVRLIM 246
QY 272 LRDRDSGPKV-TGNAALTLRNFCWAKQKLNKVS DKHPEYWDTAIFTRQDLGCGATTCDTL 330
DQ 247 VGYRQSLSLIERGNPSRSLQVCRWAHSQQRDPSHAEHHDHVVFTRQDF-GPSGQ-- 303
QY 331 GMADVGTMCDPKRCSCSVIEDDGLPSAFTTAHELGHVFNMPHD---NVKYCEEVFGKLRAN 387
DQ 304 GYAPVTGMCHPLRSCALNHEDGFSFAFVIAHETGHVLMGHDGQNGCADETSLSGS---- 359
QY 388 HMASPTLIQIDRANPWSACSAAITDFLDSCGDCLLDPSKPI-SLPEDLPGASYTLQ 446
DQ 360 -VMAPLVQAARFHRHWSRCSKLESLRPLS--YDCLDDPDPDPANPQPPPELPGINYSME 416
QY 447 QCELAFGVGSKPC---PYMOYCTKLWCTGKAKGMQVQOTRHPFWADGTSCGEGKLC LKA 503
DQ 417 QCRFDGSGYQTCIAFRTEPCQKLMCS-HPDNYPYFCKTKKGPPLDGTGECAPKWCFKGH 475
QY 504 CVERHNLNKHVRDGSWAKWDYPGCSRTCGGGVQLARRQCTNPTPANGKYCEGVVRKYR 563
DQ 476 CIWKSPEQTYGQDGGWSSWTKFGSCRS CGGVRSRSCNNPSPAYGGRPCGLGPMFEYQ 535
QY 564 SCNLEPCPSSASGKSFREEQ 583
DQ 536 VCNSECEPGTY--EDFRAQ 553
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:25 ; Search time 28.3707 Seconds
(without alignments)
2738.218 Million cell updates/sec

Title: US-10-009-332-1-copy_1_583

Perfect score: 3163

Sequence: 1 MLLGLITLAFAGTAGGFE.....SCNLEPCPSASGKSFRRQ 583

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3163	100.0	950	22	Human metalloprotease
2	3155	99.7	928	23	Human metalloprotease #2
3	3155	99.7	950	23	Human protease PRT
4	3155	99.7	952	23	Rat metalloprotease #1
5	2497.5	79.0	505	21	Human metalloprotease #3
6	1913	60.5	367	23	Human metalloprotease #1
7	1767	55.9	381	21	Human metalloprotease #1
8	1669	52.8	321	23	Human metalloprotease #1
9	1562.5	49.4	896	21	Human metalloprotease #1
10	1560.5	49.3	950	20	Human metalloprotease #1

11	1560.5	49.3	950	22	Human metalloprotease #1
12	1560.5	49.3	950	22	Human metalloprotease #1
13	1560.5	49.3	967	19	Human metalloprotease #1
14	1560.5	49.3	968	22	Human metalloprotease #1
15	1559.5	49.3	967	20	Human metalloprotease #1
16	1538.5	48.6	967	20	Human metalloprotease #1
17	1536	48.6	950	21	Human metalloprotease #1
18	1362	43.1	890	20	Human metalloprotease #1
19	1362	43.1	890	22	Human metalloprotease #1
20	1359	43.0	889	22	Human metalloprotease #1
21	1351	42.7	727	20	Human metalloprotease #1
22	1338.5	42.3	837	20	Human metalloprotease #1
23	1337.5	42.3	837	21	Human metalloprotease #1
24	1337.5	42.3	837	22	Human metalloprotease #1
25	1337.5	42.3	837	22	Human metalloprotease #1
26	1336.5	42.3	837	22	Human metalloprotease #1
27	1336.5	42.3	840	21	Human metalloprotease #1
28	1329.5	42.0	905	22	Human metalloprotease #1
29	1240	39.2	930	22	Human metalloprotease #1
30	1232.5	39.0	947	22	Human metalloprotease #1
31	1232.5	39.0	1629	23	Human metalloprotease #1
32	1232.5	39.0	1629	23	Human metalloprotease #1
33	1232.5	39.0	1916	23	Human metalloprotease #1
34	1232.5	39.0	1935	23	Human metalloprotease #1
35	1229.5	38.9	1602	23	Human metalloprotease #1
36	1226.5	38.8	1073	21	Human metalloprotease #1
37	1217.5	38.5	870	21	Human metalloprotease #1
38	1212.5	38.3	930	20	Human metalloprotease #1
39	1212	38.3	1934	22	Human metalloprotease #1
40	1205	38.1	929	21	Human metalloprotease #1
41	1160	36.7	1505	23	Human metalloprotease #1
42	1153.5	36.5	625	23	Human metalloprotease #1
43	1149.5	36.3	1907	23	Human metalloprotease #1
44	1142.5	36.1	680	21	Human metalloprotease #1
45	1121	35.4	1882	22	Human metalloprotease #1

ALIGNMENTS

RESULT 1

AAAG62299 standard; protein; 950 AA.

XX AAAG62299;

AC 23-AUG-2001 (first entry)

DT Human metalloprotease MDT56 protein.

DE Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
KW osteopath; antiarthritic.

OS Homo sapiens.

XX WO200134785-A1.

XX 17-MAY-2001.

XX 10-NOV-2000; 2000WO-JP07917.

XX 11-NOV-1999; 99JP-0321740.

XX 16-MAY-2000; 2000JP-0144020.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

XX (KAZU-) KAZUSA DNA RES INST.

XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

XX N-PSDB; AAH41003.

PT Metalloprotease with aggrecanase activity for treating joint diseases

PT especially osteoarthritis -

PS Claim 1; Page 56-60; 85pp; Japanese.

XX This invention relates to a metalloprotease with aggreganase activity.

CC The invention includes protein and DNA sequences of the metalloprotease,

CC vectors containing the DNA, host cells transformed by the vectors, and

CC antibodies directed against the metalloprotease. The antibodies, protein

CC and DNA sequences can be used in the treatment and prevention of joint

CC diseases, particularly osteoarthritis. The treatment may result in

CC osteopathic and antiarthritic activity. The present sequence represents

CC the metalloprotease of the invention termed MDTs6.

XX

XX Sequence 950 AA;

XX

Query Match 100.0%; Score 3163; DB 22; Length 950;

Best Local Similarity 100.0%; Pred. No. 7.4e-255;

Matches 583; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDSDGGLIFQITAF 60

DB 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDSDGGLIFQITAF 60

QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAVSLC 120

DB 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAVSLC 120

QY 121 GGLRGAFYGAEVVISPLPNASAPAAQNSQGAHLQRRGVGGPSGDTSCRGVASGW 180

DB 121 GGLRGAFYGAEVVISPLPNASAPAAQNSQGAHLQRRGVGGPSGDTSCRGVASGW 180

QY 181 NPAILRALDPKPRRAGFESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 240

DB 181 NPAILRALDPKPRRAGFESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 240

QY 241 YLLTLLATAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALTIRNFCWOKKLN 300

DB 241 YLLTLLATAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALTIRNFCWOKKLN 300

QY 301 KVSQKHEPYWDATLFTFRQDLGATTCDTLGMADVGTMCDPKRSVIEDDGLPSAFTTA 360

DB 301 KVSQKHEPYWDATLFTFRQDLGATTCDTLGMADVGTMCDPKRSVIEDDGLPSAFTTA 360

QY 361 HELGHVFNMPHDNKKVCEVFGKLRANHMSPTLIQIDRANPWSACSAAIITDLSGSHG 420

DB 361 HELGHVFNMPHDNKKVCEVFGKLRANHMSPTLIQIDRANPWSACSAAIITDLSGSHG 420

QY 421 DCLLDQPSKPISLPEDLPGASYTILSQOCELAFGVGSKPCPYMQYCTKLWCTGKAKGMVC 480

DB 421 DCLLDQPSKPISLPEDLPGASYTILSQOCELAFGVGSKPCPYMQYCTKLWCTGKAKGMVC 480

QY 481 QTRHFPWADGTSCEGKLCCLKGACVERHNLNKHVDGWSWAKWDYPGCSRTCGGGVQLAR 540

DB 481 QTRHFPWADGTSCEGKLCCLKGACVERHNLNKHVDGWSWAKWDYPGCSRTCGGGVQLAR 540

QY 541 RQCTNPTPANGKCYCEGVVRYKRNLEPCPSSASGKSFRREQ 583

DB 541 RQCTNPTPANGKCYCEGVVRYKRNLEPCPSSASGKSFRREQ 583

RESULT 2

AAU72899

XX AAU72899 standard; Protein: 928 AA.

XX

XX AAU72899;

XX

XX 26-FEB-2002 (first entry)

XX Human metalloprotease partial protein sequence #11.

XX

XX Human: protease; PCR primer; cytostatic; immunomodulator; cardiant;

XX vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;

XX hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;

KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;

KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;

KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;

KW immune-related disease; cardiovascular disease; neuronal disease;

KW migraine; sexual dysfunction; mood disorder; attention disorder;

KW cognition disorder; hypotension; hypertension; psychotic disorder;

KW dyskinesia; metabolic disorder; inflammatory disorder.

OS Homo sapiens.

XX WO200183782-A2.

PN 08-NOV-2001.

XX

XX 04-MAY-2001; 2001WO-US14431.

XX

XX 04-MAY-2000; 2000US-201879P.

XX

XX (SUGE-) SUGEN INC.

XX Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;

PI Payne V;

XX

DR WPI: 2002-041502/05.

DR N-PSDB; AAS97182.

XX

PT Novel protease polypeptide useful for screening for substances that may

PT be used to treat, e.g., cancers, immune-related diseases,

PT cardiovascular disease, migraine, pain, psychotic and inflammatory

PT disorders -

XX

PS Claim 28; Figure 2G; 232pp; English.

XX

CC The invention relates to an isolated, enriched, or purified protease

CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to

CC screen for substances (S) that may modulate its activity. Administering

CC S (which modulates protease activity in vitro) may be used to treat a

CC disease or disorder selected from cancers (e.g., of tissues, of blood or

CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,

CC brain, ovarian, bladder or kidney), immune-related diseases and

CC disorders, cardiovascular disease, brain or neuronal-associated diseases

CC (e.g., central or peripheral nervous system diseases, migraine, pain,

CC sexual dysfunction, mood disorders, attention disorders, cognition

CC disorders, hypotension, hypertension, psychotic disorders, neurological

CC disorders and dyskinesias), metabolic disorders and inflammatory

CC disorders. (I) may also be useful as a diagnostic tool for a disease or

CC disorder such as those above. AAU72876-AAU72910 represent human

CC protease amino acid sequences of the invention.

XX

XX Sequence 928 AA;

XX

Query Match 99.7%; Score 3155; DB 23; Length 928;

Best Local Similarity 99.8%; Pred. No. 3.3e-254;

Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDSDGGLIFQITAF 60

DB 26 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDSDGGLIFQITAF 85

QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAVSLC 120

DB 86 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAVSLC 145

QY 121 GGLRGAFYGAEVVISPLPNASAPAAQNSQGAHLQRRGVGGPSGDTSCRGVASGW 180

DB 146 GGLRGAFYGAEVVISPLPNASAPAAQNSQGAHLQRRGVGGPSGDTSCRGVASGW 205

QY 181 NPAILRALDPKPRRAGFESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 240

DB 206 NPAILRALDPKPRRAGFESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 265

QY 241 YLLTLLATAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALTIRNFCWOKKLN 300

Db 266 YLLTLATAARLYRHPSILNPINIVVKKVLLLRDRDSGPKVTGNAALTLRNFCWQKKLN 325
 QY 301 KVSDEKHEPYWDTAIILFTRODLGGATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 360
 Db 326 KVSDEKHEPYWDTAIILFTRODLGGATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 385
 QY 361 HELGHVFNPHDNVNVKCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDGSHG 420
 Db 386 HELGHVFNPHDNVNVKCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDGSHG 445
 QY 421 DCLLDQPSKPISLPEDLPASVYTLSSQCELAFGVGSKPCPYMOYCTKLWCTGKAGQMVC 480
 Db 446 DCLLDQPSKPISLPEDLPASVYTLSSQCELAFGVGSKPCPYMOYCTKLWCTGKAGQMVC 505
 QY 481 QTRHFPWADGTSCEGKLCCKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGGVQLAR 540
 Db 506 QTRHFPWADGTSCEGKLCCKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGGVQLAR 565
 QY 541 RQCTNPTPANGKYCEGVVRYKRSNLEPCPSSASGKSFREQ 583
 Db 566 RQCTNPTPANGKYCEGVVRYKRSNLEPCPSSASGKSFREQ 608

RESULT 3
 AAE22541
 ID AAE22541 standard; Protein; 950 AA.

XX AAE22541;
 XX

DT 26-JUL-2002 (first entry)

XX Human protease #2.

XX Human; novel human protein; NHP; protease; biological disorder; obesity;
 KW high blood pressure; arthritis; connective tissue disorder; infertility;
 KW gene therapy; enzyme.

XX Homo sapiens.

XX WO200226949-A2.

XX 04-APR-2002.

XX 27-SEP-2001; 2001WO-US30350.

XX 29-SEP-2000; 2000US-236689P.

XX (LEXI-) LEXICON GENETICS INC.

XX Friddle CJ, Hilbun E;

XX WPI; 2002-372123/40.

DR N-PSDB; AAD35569.

XX Novel nucleic acid encoding a human protease, useful as a hybridization
 PT probe for screening libraries and assessing gene expression patterns -
 PT Claim 6; Page 36-38; 41pp; English.

XX The present sequence is novel human protein (NHP), human protease.
 CC NHPs share structural similarity with animal proteases particularly
 CC zinc metalloproteases. Sequences of the invention are useful in
 CC therapeutic, diagnostic and pharmacogenomic applications. NHP
 CC polynucleotides are used as hybridisation probes for screening
 CC libraries and assessing gene expression patterns. They can also be
 CC used for treating related biological disorders such as obesity, high
 CC blood pressure, arthritis, connective tissue disorders and infertility.
 CC They are also used in gene therapy.

XX Sequence 950 AA;

Query Match 99.7%; Score 3155; DB 23; Length 950;
 Best Local Similarity 99.8%; Pred. No. 3.4e-254.

Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MLLGILTLAFAGRTAGGPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFOITAF 60
 Db 1 MLLGILTLAFAGRTAGGPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFOITAF 60
 QY 61 QEDFYHLHLPDPAFLAPAFSTEHLGVLPLQGLTGGSSDLRRRCFYSGDVNAEPDSFAVSLC 120
 Db 61 QEDFYHLHLPDPAFLAPAFSTEHLGVLPLQGLTGGSSDLRRRCFYSGDVNAEPDSFAVSLC 120
 QY 121 GGLRGAFGVRGABYVLSPLNASAPAAQRNSQGAHLQLRRGVPGGPGSDPTSRCGVASGW 180
 Db 121 GGLRGAFGVRGABYVLSPLNASAPAAQRNSQGAHLQLRRGVPGGPGSDPTSRCGVASGW 180
 QY 181 NPAILRALDPYKPRRAGFGESRRRSRAKRFVSIPIRYVETLVVADESMMVKFHGADLEH 240
 Db 181 NPAILRALDPYKPRRAGFGESRRRSRAKRFVSIPIRYVETLVVADESMMVKFHGADLEH 240
 QY 241 YLLTLATAARLYRHPSILNPINIVVKKVLLLRDRDSGPKVTGNAALTLRNFCWQKKLN 300
 Db 241 YLLTLATAARLYRHPSILNPINIVVKKVLLLRDRDSGPKVTGNAALTLRNFCWQKKLN 300
 QY 301 KVSDEKHEPYWDTAIILFTRODLGGATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 360
 Db 301 KVSDEKHEPYWDTAIILFTRODLGGATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 360
 QY 361 HELGHVFNPHDNVNVKCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDGSHG 420
 Db 361 HELGHVFNPHDNVNVKCEVEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDGSHG 420
 QY 421 DCLLDQPSKPISLPEDLPASVYTLSSQCELAFGVGSKPCPYMOYCTKLWCTGKAGQMVC 480
 Db 421 DCLLDQPSKPISLPEDLPASVYTLSSQCELAFGVGSKPCPYMOYCTKLWCTGKAGQMVC 480
 QY 481 QTRHFPWADGTSCEGKLCCKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGGVQLAR 540
 Db 481 QTRHFPWADGTSCEGKLCCKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGGVQLAR 540
 QY 541 RQCTNPTPANGKYCEGVVRYKRSNLEPCPSSASGKSFREQ 583
 Db 541 RQCTNPTPANGKYCEGVVRYKRSNLEPCPSSASGKSFREQ 583

RESULT 4

AAU74751

ID AAU74751 standard; Protein; 952 AA.

XX AAU74751;
 AC AAU74751;

DT 09-APR-2002 (first entry)

XX Human protease PRTS-11 protein sequence.

XX Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
 KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
 KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
 KW cell proliferative disorder; developmental disorder; epilepsy;
 KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
 KW reproductive disorder; endometriosis.

XX Homo sapiens.

XX WO200198468-A2.

XX 27-DEC-2001.

XX 13-JUN-2001; 2001WO-US19178.

XX 16-JUN-2000; 2000US-212336P.

XX 22-JUN-2000; 2000US-213955P.

XX 29-JUN-2000; 2000US-213396P.

XX 07-JUL-2000; 2000US-216821P.

XX 14-JUL-2000; 2000US-218946P.

XX (INCY-) INCYTE GENOMICS INC.
 XX Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
 PI Belegeane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
 PI Wallia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
 PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
 PI Kalllick DA;
 XX WPI: 2002-090437/12.
 DR N-PSDB; ABK12894.
 XX
 XX Twenty one human proteases (referred to as PRS-1 to PRS-21), useful
 PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
 PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
 PT proliferative (e.g. cancer) disorders -
 XX Claim 1; Page 144-146; 177pp; English.
 XX
 XX The present invention relates to twenty one new human proteases,
 CC referred to as PRS-1 to PRS-21. The PRS polynucleotides and
 CC polypeptides of the invention are useful in the diagnosis, treatment and
 CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
 CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
 CC myocardial infarction, autoimmune/inflammatory e.g. acquired
 CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
 CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
 CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
 CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
 CC endometriosis disorders. Numerous other examples of each disorder are
 CC given in the specification. The present protein sequence represents
 CC the human protease PRS-11 protein of the invention.
 XX
 XX Sequence 952 AA;
 SQ
 Query Match. 99.7%; Score 3155; DB 23; Length 952;
 Best Local Similarity 99.8%; Pred. No. 3.4e-254;
 Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MLLGLTLTAFAGTAGGFEVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
 DB 1 MLLGLTLTAFAGTAGGFEVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
 QY 61 QEDFYLHLTPDAQPLAFSTHGLVPLQGLTGSSDLRCFCYSGDVNAEPDSFAAVSLC 120
 DB 61 QEDFYLHLTPDAQPLAFSTHGLVPLQGLTGSSDLRCFCYSGDVNAEPDSFAAVSLC 120
 QY 121 GGLRGAFYRGAEYVISPPLNASAPAAQRNSOGAHLQRRGVPGSPGDPTRSCGVASGW 180
 DB 121 GGLRGAFYRGAEYVISPPLNASAPAAQRNSOGAHLQRRGVPGSPGDPTRSCGVASGW 180
 QY 181 NPAILRALDPYKPRRAGFGESRRSRSSRAKRFVSTPRYVETLVADESVMKFGADLEH 240
 DB 181 NPAILRALDPYKPRRAGFGESRRSRSSRAKRFVSTPRYVETLVADESVMKFGADLEH 240
 QY 241 YLLTLLAARLYRHPILPINIVVVKVLLLRDRDSGPKVTGNAALTFRNCAWOKKLN 300
 DB 241 YLLTLLAARLYRHPILPINIVVVKVLLLRDRDSGPKVTGNAALTFRNCAWOKKLN 300
 QY 301 KVSQKHEPYWDITLFTQDLCCGATTCDTLGMADVGTMCDPKRSVIEDDGLPSAFTTA 360
 DB 301 KVSQKHEPYWDITLFTQDLCCGATTCDTLGMADVGTMCDPKRSVIEDDGLPSAFTTA 360
 QY 361 HELGHVFNMPHDNFKVCEEVFGKLRANHMSPILQIDRANPWSACSAAIITFDLSGGH 420
 DB 361 HELGHVFNMPHDNFKVCEEVFGKLRANHMSPILQIDRANPWSACSAAIITFDLSGGH 420
 QY 421 DCLLDQFSKPISTLPEDLPGASYTILSQOCELAFGVSKPCPYMCTKLWCTGKAKGMVC 480
 DB 421 DCLLDQFSKPISTLPEDLPGASYTILSQOCELAFGVSKPCPYMCTKLWCTGKAKGMVC 480
 QY 481 QTRHFPWADGTSCEGKGLCLKGACVERHNLNKHRVDGSAKWADPYGPCSRTCGGVOLAR 540

DB 481 QTRHFPWADGTSCEGKGLCLKGACVERHNLNKHRVDGSAKWADPYGPCSRTCGGVOLAR 540
 QY 541 RQCTNPFPANGKYGCEGVVYKRSCLNLEPCPSSASGKSFREEQ 583
 DB 541 RQCTNPFPANGKYGCEGVVYKRSCLNLEPCPSSASGKSFREEQ 583
 RESULT 5
 AAB21257
 ID AAB21257 standard; Protein; 505 AA.
 XX
 AC AAB21257;
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Rat metalloproteinase ADAMTS-5.
 XX
 KW Rat; ADAMTS-5; metalloproteinase; ADAM;
 KW a disintegrin and metalloproteinase domain; thrombospondin domain;
 KW vaccine; neurotropic; neuroprotective; antiparkinsonian;
 KW cerebroprotective; cytoskeletal; antiarthritic; immunosuppressive;
 KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
 KW autoimmune disease; brain tumour; brain injury.
 OS
 XX Rattus norvegicus.
 XX
 PN WO200053774-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US06237.
 XX
 PR 08-MAR-1999; 99US-0264585.
 XX
 PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
 XX
 PI Kelner GS, Clark M, Maki RA;
 XX
 DR WPI: 2000-594326/56.
 DR N-PSDB; AAA95827.
 XX
 PT Polynucleotide encoding novel members of a disintegrin,
 PT metalloproteinase and thrombospondin domain protein family used to
 PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
 XX
 PS Claim 12; Fig 14; 129pp; English.
 XX
 CC The present sequence is rat metalloproteinase ADAMTS-5. The
 CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
 CC and Metalloproteinase Domain) family. Members of the ADAMTS family
 CC contain a thrombospondin domain in addition to the disintegrin and
 CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
 CC useful for the manufacture of medicaments for treating conditions
 CC associated with neuroinflammation and/or neurodegeneration, such as
 CC Alzheimer's disease, Parkinson's disease and stroke. They are also
 CC useful for treating conditions associated with cell proliferation, cell
 CC migration, inflammation and/or angiogenesis, such as cancer, arthritis,
 CC and autoimmune diseases. They can be used to treat patients afflicted
 CC with an invasive tumour, a brain tumour or brain injury.
 XX
 SQ Sequence 505 AA;

Query Match 79.0%; Score 2497.5; DB 21; Length 505;
 Best Local Similarity 93.5%; Pred. No. 1e-199;
 Matches 462; Conservative 8; Mismatches 21; Indels 3; Gaps 2;
 QY 89 QGLTGSSDLRCFCYSGDVNAEPDSFAAVSLCGLRGAFYRGAEYVISPPLNASAPAAQ 148
 DB 13 QRLTGSSDLRCFCYSGYVNAEPDSFAAVSLCGLRGAFYRGAEYVISPPLNTSAPAAQ 72
 QY 149 RNSOGAHLQRRGVPGSPGDPTRSCGVASGNPAILRALDPYKPRRAGFGESRRSRSSG 208
 DB 73 RNSOGAHLQRRGVPGSPGDPTRSCGVASGNPAILRALDPYKPRRAGFGESRRSRSSG 132

QY 209 RAKREVSIPIRVETLVVADESMVFGADLEHLLTLLATAARLYRHPSILNPNIVVVK 268
 DB 133 RAKREVSIPIRVETLVVADESMVFGADLEHLLTLLATAARLYRHPSILNPNIVVVK 192
 QY 269 VLLLRDRSGKVTGNAALTIRNFCAMQKLNKYSKHPYWDTAILFTRODLGATTC 328
 DB 193 VLLLRDRSGKVTGNAALTIRNFCAMQKLNKYSKHPYWDTAILFTRODLGATTC 252
 QY 329 TLGMADVCTMCDPKRSCSVIEDDGLPSAFTTAHELGHVENPHDNVCEVFGKLRANH 388
 DB 253 TLGMADVCTMCDPKRSCSVIEDDGLPSAFTTAHELGHVENPHDNVCEVFGKLRANH 312
 QY 389 MMSPTLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKIPSLPEDLPASYSLSQOC 448
 DB 313 MMSPTLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKIPSLPEDLPASYSLSQOC 372
 QY 449 ELAFVGSGKPCPYMOYCTKLACTGKAKQMVCOYTRHFPWADGTSCEGKFLGACVYERH 508
 DB 373 ELAFVGSGKPCPYMOYCTKLACTGKAKQMVCOYTRHFPWADGTSCEGKFLGACVYERH 432
 QY 509 NLNHRVDGSKWADPYGPCSRTCGGQVQLARR--OCTNPTPANGKYCEGVRVKYRSCN 566
 DB 433 NPKNYVDGPNKWEPIGPCSRTCGGQVQLARRQVQATLPLP-TGGKYCEGVRVKYRSCN 491
 QY 567 LEPCPSSASGKSFR 580
 DB 492 LEPCPSSASGKSFR 505

RESULT 6
 AAE22542
 ID AAE22542 standard; Protein; 367 AA.
 AC AAE22542;
 DT 26-JUL-2002 (first entry)
 DE Human protease #3.
 KW Human; novel human protein; NHP; protease; biological disorder; obesity;
 KW high blood pressure; arthritis; connective tissue disorder; infertility;
 KW gene therapy; enzyme.
 OS Homo sapiens.
 FN WO200226949-A2.
 PD 04-APR-2002.
 PF 27-SEP-2001; 2001WO-US30350.
 PR 29-SEP-2000; 2000US-236689P.
 PA (LEXI-) LEXICON GENETICS INC.
 PI Friidde CJ, Hilbun E;
 DR WPI; 2002-372123/40.
 DR N-PSDB; AAD35570.
 PT Novel nucleic acid encoding a human protease, useful as a hybridization
 PT probe for screening libraries and assessing gene expression patterns -
 XX Claim 7; Page 39-40; 41pp; English.
 XX The present sequence is novel human protein (NHP), human protease.
 CC NHPs share structural similarity with animal proteases particularly
 CC zinc metalloproteases. Sequences of the invention are useful in
 CC therapeutic, diagnostic and pharmacogenomic applications. NHP
 CC polynucleotides are used as hybridisation probes for screening
 CC libraries and assessing gene expression patterns. They can also be
 CC used for treating related biological disorders such as obesity, high

CC blood pressure, arthritis, connective tissue disorders and infertility.
 CC They are also used in gene therapy.
 SQ Sequence 367 AA;
 Query Match 60.5%; Score 1913; DB 23; Length 367;
 Best Local Similarity 99.7%; Pred. No. 4e-151;
 Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MLLLGILTLTAFAGTAGGPEPEVVPVIRLDPDINGRYYWRGPDSDGQGLIFQITAF 60
 DB 1 MLLLGILTLTAFAGTAGGPEPEVVPVIRLDPDINGRYYWRGPDSDGQGLIFQITAF 60
 QY 61 QEDFYLHLFPDQAFAPAFSTEHLGVLPLQGLTGGSSDLRRCFVSGDVNAEPDSFAVSLC 120
 DB 61 QEDFYLHLFPDQAFAPAFSTEHLGVLPLQGLTGGSSDLRRCFVSGDVNAEPDSFAVSLC 120
 QY 121 GGLRGAFYRGAEYVISPPLNASAPAAQNSQGAHLQRRGVPGGSPGSDPTSRGCVASGW 180
 DB 121 GGLRGAFYRGAEYVISPPLNASAPAAQNSQGAHLQRRGVPGGSPGSDPTSRGCVASGW 180
 QY 181 NPAILRALDPYKPRRAGFGESESRSSRRSGRAKRFVSPRIVVELVVADESMVFGADLEH 240
 DB 181 NPAILRALDPYKPRRAGFGESESRSSRRSGRAKRFVSPRIVVELVVADESMVFGADLEH 240
 QY 241 YLLTLTATAARLYRHPSILNPNIVVVKVLLLRDRSGPKVTGNAALTIRNFCAMQKLN 300
 DB 241 YLLTLTATAARLYRHPSILNPNIVVVKVLLLRDRSGPKVTGNAALTIRNFCAMQKLN 300
 QY 301 KYSDKHPEYWDTAILFTRODLGATTCDTLGMADVCTMCDPKRSCSVIEDDGLPSAFTTA 360
 DB 301 KYSDKHPEYWDTAILFTRODLGATTCDTLGMADVCTMCDPKRSCSVIEDDGLPSAFTTA 360
 QY 361 HELG 364
 DB 361 HELG 364

RESULT 7
 AAB21261
 ID AAB21261 standard; Protein; 381 AA.
 XX AAB21261;
 DT 23-FEB-2001 (first entry)
 DE Human metalloproteinase ADAMTS-5.
 KW Human; ADAMTS-5; metalloproteinase; ADAM;
 KW a disintegrin and metalloproteinase domain; thrombospondin domain;
 KW vaccine; neutropic; neuroprotective; antiparkinsonian;
 KW cerebroprotective; cytostatic; antiarthritis; immunosuppressive;
 KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
 KW autoimmune disease; brain tumour; brain injury.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FH Misc-difference 196 /note= "Xaa- any amino acid"
 FT Misc-difference 268 /note= "Xaa- any amino acid"
 FT Misc-difference 304 /note= "Xaa- any amino acid"
 FT Misc-difference 308 /note= "Xaa- any amino acid"
 FT /note= "Xaa- any amino acid"
 XX WO200053774-A2.
 XX 14-SEP-2000.
 XX 08-MAR-2000; 2000WO-US06237.
 XX

```

PR 08-MAR-1999; 99US-0264585.
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX Kelner GS, Clark M, Maki RA;
XX WPI; 2000-594326/56.
DR N-PSDB; AAA95831.
XX Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Claim 12; Fig 23; 129pp; English.
XX The present sequence is human metalloproteinase ADAMTS-5. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.
XX Sequence 381 AA;
SQ
Query Match 55.9%; Score 1767; DB 21; Length 381;
Best Local Similarity 93.3%; Pred. No. 6.3e-139;
Matches 322; Conservative 3; Mismatches 18; Indels 2; Gaps 2;
QY 240 HYLLTLTAARLYRHPSTLNINIVVVKVLLLRDRDSCPKVTGNAAFLRNFCWQKKL 299
DB 2 HYRRAAARAG-IFKHPSILNPNINIVVVKVLLLRDRDSCPKVTGNAAFLRNFCWQKKL 60
QY 300 NKYSKDHPEYWDTAIFLTRODLGATTCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTT 359
DB 61 NKYSKDHPEYWDTAIFLTRODLGATTCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTT 120
QY 360 AHELGHVFNPHDNVKEEVFGKLRANHMSPSTLIQIDRANPWSACSAIITDFLDSGH 419
DB 121 AHELGHVFNPHDNVKEEVFGKLRANHMSPSTLIQIDRANPWSACSAIITDFLDSGH 180
QY 420 GDCLLDQPSKPISLPDLPGASVTLTQQCELAFGVGSKPCPYMOYCTKLWCTGKAKGMV 479
DB 181 GDCLLDQPSKPIELPDLPGASVTLTQQCELAFGVGSKPCPYMOYCTKLWCTGKAKGMV 240
QY 480 CQTRHFPWADTSCGEGKCLKLGACVYERHNLNKHRYVDGSKAWDYPGCPSCRTGGGVQLA 539
DB 241 CQTRHFPWADTSCGEGKCLKLGACVEXHNLNKHRYVDGSKAWDYPGCPSCRTGGGVQLA 300
QY 540 RROCTNPTP-ANGGKVCYGVYKRYSCNLEPCPSSASGKSFRREQ 583
DB 301 RROXHOXPPLTGGKYCEGVYKRYSCNLEPCPSSASGKSFRREQ 345
RESULT 8
AAE22540
ID AAE22540 standard; Protein; 321 AA.
XX
AC AAE22540;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #1.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX

```

```

OS Homo sapiens.
XX WO200226949-A2.
XX 04-APR-2002.
XX 27-SEP-2001; 2001WO-US30350.
XX 29-SEP-2000; 2000US-236689P.
XX (LEXI-) LEXICON GENETICS INC.
XX Friddle CJ, Hilbun E;
XX WPI; 2002-372123/40.
XX N-PSDB; AAD35568.
XX Novel nucleic acid encoding a human protease, useful as a hybridization
XX probe for screening libraries and assessing gene expression patterns -
XX Claim 5; Page 34-35; 41pp; English.
XX The present sequence is novel human protein (NHP), human protease.
XX NHPs share structural similarity with animal proteases particularly
XX zinc metalloproteases. Sequences of the invention are useful in
XX therapeutic, diagnostic and pharmacogenomic applications. NHP
XX polynucleotides are used as hybridisation probes for screening
XX libraries and assessing gene expression patterns. They can also be
XX used for treating related biological disorders such as obesity, high
XX blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy.
XX Sequence 321 AA;
SQ
Query Match 52.8%; Score 1669; DB 23; Length 321;
Best Local Similarity 99.4%; Pred. No. 7.5e-131;
Matches 318; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLLGILTLAFAGTAGGFEPEYVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
DB 1 MLLGILTLAFAGTAGGFEPEYVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
QY 61 QEDFYLHTLPAQFLAPAFSTHGLGVPLQGLTGSSDLRRCFYSGDYNAPDSFAAVSLC 120
DB 61 QEDFYLHTLPAQFLAPAFSTHGLGVPLQGLTGSSDLRRCFYSGDYNAPDSFAAVSLC 120
QY 121 GGLGAFGYRGAEEYVISPPLNASAPAAQNSQGAHLQLQRRGVPGSGDPTSRGCVASGW 180
DB 121 GGLGAFGYRGAEEYVISPPLNASAPAAQNSQGAHLQLQRRGVPGSGDPTSRGCVASGW 180
QY 181 NPAILRALDPYKPRRAGGESRRSRRAKRFVSIPIRYVETLVVADESVMKFGADLEH 240
DB 181 NPAILRALDPYKPRRAGGESRRSRRAKRFVSIPIRYVETLVVADESVMKFGADLEH 240
QY 241 YLLTLLATAARLYRHPSTLNINIVVVKVLLLRDRDSCPKVTGNAAFLRNFCWQKKLN 300
DB 241 YLLTLLATAARLYRHPSTLNINIVVVKVLLLRDRDSCPKVTGNAAFLRNFCWQKKLN 300
QY 301 KVS DKHPEYWDTAIFLTROD 320
DB 301 KVS DKHPEYWDTAIFLTROE 320
RESULT 9
AAB21265
ID AAB21265 standard; Protein; 896 AA.
XX
AC AAB21265;
XX
DT 23-FEB-2001 (first entry)
XX
DE Mouse metalloproteinase ADAMTS-1.
XX

```


CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulomas, hypertrophic scars, nonunion fractures,
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
 CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH can also be used in birth control. METH can also
 CC be used in diagnostic methods for the prognosis of cancer. The present
 CC sequence is a protein isolated in the present invention.
 XX
 SQ Sequence 968 AA;
 Query Match 49.3%; Score 1560.5; DB 22; Length 968;
 Best Local Similarity 51.6%; Pred. No. 3.8e-121;
 Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;
 QY 1 MLLLGILTLAFAGTAGGFEPEREVVPRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
 Db 37 LLLLLAALLAVSDALGRSEDEELVVP-ELE-----RAP---GHGTTTLRLHAF 82
 QY 61 QEDFYHLTPDQAFLAPAFSTEHLG-----VPLQGLTGGSDLRRCFYSGDVNAEPDSF 114
 Db 83 DQQLDLRLPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFSYGTNGDPSSA 137
 QY 115 AAVSLCGGLRGAFGYGAEYVISPLPNAS---APAAQRNSOGA----HLLQ---RRGVPGG 165
 Db 138 AALSCEGVRGAFYLLGEAYFIQPLPAASERLATAAGEKPPAPLQFHLRLNRNGDVG 197
 QY 166 PSG-----DPTSRC-----GVASG--WNPAIRLALDPYKPRRAGFSGSRSSRRS 207
 Db 198 TCGVVDDEPRPTGKAETDEDEGTEGEGPQWS-----PQDPALQGVGP-TGTGS 248
 QY 208 GRAKRFVSIPRYVETLVVADESMVKFHGADLEHYLLTLLATAARLYRHPISILNINIVV 267
 Db 249 IRKKRFVSSHRYVETMLVADQSMAEFHGSLGKHYLLTLFSAARLYKHPISIRNSVLVV 308
 QY 268 KVLRLRDSGPKVTGNAALTNRNFCWQKLNKVSQKHPEYWDTAIFTRQDLGGATTC 327
 Db 309 KILVHDEQKGPEVTSNAALTNRNFCWQKLNKVSQKHPEYWDTAIFTRQDLGGATTC 368
 QY 328 DTLMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCCEVEFGKLRAN 387
 Db 369 DTLMADVGTVCDFSRCSVIEDDGLQAAFTTAHELGHVFNMPHDNVKVCCEVEFGKLRAN 428
 QY 388 HMSNPTLIQIDRANPWSACSAAIITDFLDSHGDCILLDQPSKPISLPEDLPASVYLSQ 447
 Db 429 HMASMLSLNDSQWSPSCSAYMITSFLDNGHGECIMDKPQNPQIQLPGLPFTSYDANRQ 488
 QY 448 CELAFGVGSKPCP-YMQYCTKLWCTGKAKGOMVCOTRFPWADGTSCEGKCLKGACVE 506
 Db 489 CQFTGEDSKHCPDRAASTCTLWCTGTSGLVLCQTKFHPWADGTSCEGKWCINGKCNV 548
 QY 507 RHNLNKH---RYDGSWAKWDYPGCSRTCGGVQVLAARQCTNPTTPANGKVCYCEGVRYR 563
 Db 549 KTD-RKHFTPTPHGSMGMPGDCSRTCGGVQVYTMRECDNPVPKNGKCYCEGKRYR 607
 QY 564 SCNLEPCPSASGKSFREQ 583
 Db 608 SCNLEDCPDN-NGKTFREQ 626
 RESULT 15
 AAY04142
 ID AAY04142 standard; Protein; 967 AA.
 XX
 AC AAY04142;
 XX
 DT 15-JUN-1999 (first entry)
 XX
 DE Human Tango-71 protein.

KW Human; Tango-71; Tango-73; Tango-74; Tango-76; Tango-83; diagnosis;
 KW detection.
 OS Homo sapiens.
 XX WO907850-A1.
 PN
 PD 18-FEB-1999.
 XX
 PF 06-AUG-1998; 98WO-US16502.
 XX
 PR 05-SEP-1997; 97US-0058108.
 PR 06-AUG-1997; 97US-0054966.
 XX
 XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PA
 PI Goodearl ADJ, Holtzman DA;
 XX
 XX WPI; 1999-167426/14.
 DR
 DR N-PSDB; AAX19955.
 XX
 XX New TANGO polypeptides and nucleic acids encoding them - useful as
 PT diagnostic agents and for treating disorders caused by aberrant
 PT expression of TANGO
 XX
 PS Claim 8; Fig 1; 84pp; English.
 XX
 XX The present sequence represents human Tango-71. Tango polypeptides are
 CC useful for identifying compounds which bind the polypeptide via direct
 CC binding, competition binding assays or Tango-71, -73, -74, 76 or -83-
 CC mediated signal transduction. Tango polypeptides are also useful for
 CC identifying modulating compounds by determining effect on Tango activity.
 CC Tango polypeptides and nucleic acids are useful for diagnosing diseases
 CC related to aberrant expression of Tango, and Tango polypeptides are
 CC useful for raising antibodies which can be used in diagnostic assays for
 CC detection of Tango, and also for generating anti-idiotypic antibodies for
 CC prevention and protection.
 XX
 SQ Sequence 967 AA;
 Query Match 49.3%; Score 1559.5; DB 20; Length 967;
 Best Local Similarity 51.6%; Pred. No. 4.6e-121;
 Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;
 QY 1 MLLLGILTLAFAGTAGGFEPEREVVPRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
 Db 36 LLLLLAALLAVSDALRPSSEDEELVVP-ELE-----RAP---GHGTTTLRLHAF 81
 QY 61 QEDFYHLTPDQAFLAPAFSTEHLG-----VPLQGLTGGSDLRRCFYSGDVNAEPDSF 114
 Db 82 DQQLDLRLPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFSYGTNGDPSSA 136
 QY 115 AAVSLCGGLRGAFGYGAEYVISPLPNAS---APAAQRNSOGA----HLLQ---RRGVPGG 165
 Db 137 AALSCEGVRGAFYLLGEAYFIQPLPAASERLATAAGEKPPAPLQFHLRLNRNGDVG 196
 QY 166 PSG-----DPTSRC-----GVASG--WNPAIRLALDPYKPRRAGFSGSRSSRRS 207
 Db 197 TCGVVDDEPRPTGKAETDEDEGTEGEGPQWS-----PQDPALQGVGP-TGTGS 247
 QY 208 GRAKRFVSIPRYVETLVVADESMVKFHGADLEHYLLTLLATAARLYRHPISILNINIVV 267
 Db 248 IRKKRFVSSHRYVETMLVADQSMAEFHGSLGKHYLLTLFSAARLYKHPISIRNSVLVV 307
 QY 268 KVLRLRDSGPKVTGNAALTNRNFCWQKLNKVSQKHPEYWDTAIFTRQDLGGATTC 327
 Db 308 KILVHDEQKGPEVTSNAALTNRNFCWQKLNKVSQKHPEYWDTAIFTRQDLGGATTC 367
 QY 328 DTLMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCCEVEFGKLRAN 387
 Db 368 DTLMADVGTVCDFSRCSVIEDDGLQAAFTTAHELGHVFNMPHDNVKVCCEVEFGKLRAN 427
 QY 388 HMSNPTLIQIDRANPWSACSAAIITDFLDSHGDCILLDQPSKPISLPEDLPASVYLSQ 447

```

Db 428 HMAASMLSLNLDHSPWSPCSAYMITSLDNGHGECLMDKPNPIQLPGDLPGTSDANRQ 487
QY 448 CELAFGVGSKPCP-YMOYCTKLWCTGRKAKOMVCQTRHFPWADGTSCGEGKLCIKGACVE 506
Db 488 CQFTFGEDSKHCPDAASTCTLWCTGTSGGVLYCQTKHFPWADGTSCGEGKWCINGKCVN 547
QY 507 RHNLNKH---RVDGSNAKWDYPGCSRTCGGVOLARRQCTNPTPANGGKYCEGVRYR 563
Db 548 KTD-RKHFDTPFHGSGMGMGFWGDCSRTCGGVQYTMRECDNPVKNGGKYCEGKRVYR 606
QY 564 SCNLEPCPSSASGKSFREEQ 583
Db 607 SCNLEDCPDN-NGKTFREEQ 625

```

Search completed: April 29, 2003, 17:20:33
 Job time : 30.3707 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 9.45689 Seconds
(without alignments)
1813.869 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_583
Perfect score: 3163
Sequence: 1 MLLGLTLAFAGRTAGGFE.....SCNLEPCPSASGSKSFREQ 583

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfilesi.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1560.5	49.3	967	4	US-09-130-491-2
2	1338.5	42.3	837	4	US-09-122-126B-2
3	1325.5	42.0	905	4	US-09-369-364A-9
4	1240	39.2	930	4	US-09-369-364A-2
5	1212.5	38.3	930	4	US-09-122-126B-15
6	1136	35.9	608	4	US-09-130-491-13
7	1121	35.4	1882	4	US-09-369-364A-13
8	1087.5	34.4	874	4	US-09-369-364A-15
9	932.5	29.5	997	4	US-09-369-364A-7
10	782.5	24.7	1211	4	US-09-491-522-5
11	763	24.1	1224	4	US-09-930-872-4
12	750.5	23.7	1081	4	US-09-369-364A-17
13	739	23.4	1205	4	US-09-491-522-11
14	712.5	22.5	859	4	US-09-369-364A-5
15	687	21.7	551	4	US-09-130-491-16
16	662.5	20.9	245	4	US-09-369-364A-11
17	629	19.9	518	4	US-09-369-364A-22
18	576.5	18.2	566	4	US-09-491-522-7
19	493.5	15.6	481	4	US-09-130-491-8
20	441	13.9	491	4	US-09-930-872-2
21	326	10.3	802	4	US-09-632-098-2
22	326	10.3	812	4	US-09-632-098-4
23	303	9.6	464	4	US-09-411-329C-14
24	301	9.5	462	4	US-09-411-329C-3
25	301	9.5	462	4	US-09-411-329C-17
26	292	9.2	529	2	US-08-836-442-3
27	269	8.5	616	4	US-09-608-790-1

28	268.5	8.5	621	4	US-09-026-001A-6
29	259	8.2	592	4	US-09-026-001A-14
30	256.5	8.1	621	4	US-09-026-001A-18
31	255	8.1	814	4	US-09-813-819-4
32	255	8.1	814	4	US-09-920-048-4
33	255	8.1	855	4	US-09-813-819-2
34	255	8.1	855	4	US-09-920-048-2
35	251.5	8.0	613	4	US-09-026-001A-10
36	244	7.7	335	4	US-09-152-060-64
37	238	7.5	521	4	US-09-026-001A-12
38	229.5	7.3	1170	1	US-08-313-288B-20
39	227	7.2	391	4	US-08-706-216-6
40	226	7.1	769	1	US-08-243-542-4
41	226	7.1	769	1	US-08-477-407-4
42	226	7.1	769	1	US-08-484-355-4
43	225.5	7.1	470	1	US-08-514-014-2
44	225.5	7.1	470	1	US-08-833-823-2
45	224	7.1	439	4	US-09-026-001A-8

ALIGNMENTS

RESULT 1
US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; EARLIER FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-130-491-2

Query Match 49.3%; Score 1560.5; DB 4; Length 967;
Best Local Similarity 51.6%; Pred. No. 1.5e-130;
Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;

QY 1 MLLGLTLAFAGRTAGGFEPEVEVVPIRLDPDINGRRYYWGRGPDSDGGLIFQITAF 60
Db 36 LLLAALLAVSALGRPSEDELVVP-ELE-----RAP---GHTTTLRLHAF 81

QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGSSDLRRCFYSGDYNABPDSF 114
Db 82 DQQLDLRLPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLACHFYSGTVNGPSSA 136

QY 115 AAVSLGGLRGAGYRGAEVVISPLNAS---APAAORNSQGA-----HLQ---RRGVPGG 165
Db 137 AALSCEGVRGAFYLLGEAYFIQFLPAAERLATAPEKPPAPLQFHLRLRRNQDVG 196

QY 166 PSG-----DPTSRC-----GVASG--WNPAIRALDPYKPRRAGFESRRRS 207
Db 197 TCGVDDPEPTGKAETDEDETEDEGPQWS-----PQDPAQGVGP--TGTS 247

QY 208 GRARFVSPRYVETLVVADESMVKFHDLEHLLTLLATARLYRHPISILNPINIVV 267
Db 248 IRKRFFVSSHRYVETMLVADQSNAEFGSLKHLLTFSVAARLYKHPISIRNSVLVV 307

QY 268 KVLRLDRDSGPKVTGNAALTILNFCAMOKKLNKVSUKHPEYWDTAILFTRODLGATPC 327
Db 308 KILVIHQEGPEVTSNAALTILNFCMOKHNPSPDRDAEHYDTAILFTRODLGSGTQC 367

Db 517 CGPGLCLDSCVLKEDVENRAVVDGWPWRPWGQCSRTCGGIGQFSNRECDMPMPQN 576
QY 551 GKGCEGVVRYRSCNLEPCPSSASGKSFREQ 583
Db 577 GGRFCLGERVKYQSCNTEECPP--PNGKSFREQ 607

RESULT 4
US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

Query Match 39.2%; Score 1240; DB 4; Length 930;
Best Local Similarity 44.7%; Pred. No. 5.6e-102;
Matches 242; Conservative 70; Mismatches 153; Indels 76; Gaps 7;

QY 90 GLTGSSDLRRCFYSGDVAEPDSPAAYSLCGGLRGAFYRGAEVVISPLNAPAAOR 149
Db 121 GLSASSGHRGCHFYRGTVGDSRSLAVFDLCGLDGFPAVKHARYTLKPLRGSWAEYER 180
QY 150 -----NSOGAHLQRG-----VPGPSGDPTSRGCVASGWNPAIRALDPY 191
Db 181 IYGGSSRLHYNREGFSFEALPPRASCETPASPSPG----- 217
QY 192 KPRAGFCESSRRSG-----RAKRFVSIPIRYVETLVAD 227
Db 218 -PQESPVSRSRRRSALAPOLLDSAFSPSGNAGPQWRRRRRSISRARQVELLVAD 276
QY 228 ESMVPHGADLEHYLLTLLATAARLYRHSILNPINIVVKVLLLRDRDSGPKVTYNAAL 287
Db 277 SSMARMYGRGLQHYLLTASIANRLYSHASIHRLAVVAVVVLTKDTSLEYSKNAAT 336
QY 288 TLNFCAWOKKLNKVDKHPYWDTAILFTRODLCGATCTGLMADVTGMDPKRSCSV 347
Db 337 TLKNFCWKQHQNQLGDDHEDHYDAAILFTREDLCGHSCDTGLMADVTGICSPERSCAV 396
QY 348 IEDDGLPSAFTTAHELGHVFNWPHDNVNVCEVFGKLRANHMSPTLIQIDRANPWSACS 407
Db 397 IEDDGLHAFTVAHEIGHLLGSHDDSFCEENFTGDKRLMSILTSIDASKPWSKCT 456
QY 408 AAITFDLSDGHCGLLDQPSKPISLPEDLPASVYTLSSQCELAFGVSKPCPYMOCYCTK 467
Db 457 SATITEFLDDHGNCCLLDLPKQILJLPEELPGQYDATQCNLTFGPEYVCPGMDYCAR 516
QY 468 LWCTCKAGQVQCTRHFPWADGTSCGEGKLCGLKAGCVER-----HNLNHRVDGSKWAKW 522
Db 517 LWCAVVRQGMVCLTKLPAYEGTFCGKGRVCLQCKVDKTKKKYYSTSSH--GNGSW 573
QY 523 DYPGCSRTCGGVGLAROCNTPTPANGKVCYGVVYKVSCLNEPCPSSASGKSFREE 582
Db 574 GPWGQCSRSCGGVGFAYRHCNNFAPNRSRGRTGKRAIYRSCSVTPCP--PNGKSPRHE 631
QY 583 Q 583
Db 632 Q 632

RESULT 5
US-09-122-126B-15
; Sequence 15, Application US/09122126B
; Patent No. 6451575
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 15
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-15

Query Match 38.3%; Score 1212.5; DB 4; Length 930;
Best Local Similarity 44.6%; Pred. No. 1.6e-99;
Matches 258; Conservative 71; Mismatches 180; Indels 69; Gaps 15;

QY 48 SGOGLIFQITAFQEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGSSDLR---RCFYK 104
Db 82 SGGKGVLYVAGGRFLDLDRDGSVGIAGF-----VPAGG--GTSAPMRRSHCFYR 133
QY 105 GDVNAEPDSPAAYSLCGGLRGAFYRGAEVVISPL---PNASAPAAQRNSOGA---HLL 157
Db 134 GTVDASPRSLAVFDLCGLDGFPAVKHARYTLKPLRGWAEERKRVYGDGSRALHVV 193
QY 158 QRRG-----VPGGPS-----GDPTSRCGVASGWNPAI--RALDPYK 193
Db 194 TRGFSEALPPRASCETPASTPEAHEHAPAHNSPGRALAS-----QLDQALSP--- 246
QY 194 RRAGFGS-----RSRRSRAKRFVSIPIRYVETLVADSMVKPHGLADLEHYLLTLLATAA 250
Db 247 -AGSGPQTWRRRRRS-----ISRARQVELLVADASMARLYGRGLQHYLLTLASIAN 299
QY 251 RLYRHPSILNPINIVVKVLLLRDRDGPVKTGNAALTLENFCAWOKKLNKVDKHPYV 310
Db 300 RLYSHASIHRLAVVAVVVLGDKDKEVSKNAATLKNFCWKQHQNQLGDDHEHY 359
QY 311 DTAILFTRODLCGATCTGLMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMP 370
Db 360 DAAILFTREDLCGHSCDTGLMADVTGICSPERSCAVIEDDGLHAFTVAHEIGHLLGIS 419
QY 371 HDNVKCEEVFGKLRANHMSPTLIQIDRANPWSACSAIITFDLSDGHCGLLDQPSKP 430
Db 420 HDDSKFCEETFGSTEDKRLMSILTSIDASKPWSKCTSATITEFLDDHGNCCLLDLPKQ 479
QY 431 ISLPEDLPASVYTLSSQCELAFGVSKPCPYMOCYCTKLCWCTGKAKGMVQCTRHFPWADG 490
Db 480 ILPEELPGQYDATQCNLTFGPEYVCPGMDVCARLWCAVVRQGMVCLTKLPAYEG 539
QY 491 TSCGEGKLCGLKAGCVER-----HNLNHRVDGSKWAKWDPGPCSRTCGGVGLAROCNTN 545
Db 540 TPCGKGRICLQCKVDKTKKKYYSTSSH--GNGSWSGWQCSRSCGGVGFAYRHCN 596
QY 546 PTANGKVCYGVVYRSCNLEPCPSSASGKSFREEQ 583
Db 597 PAPRNGRYCTGKRAIYRSCSLMPCP--PNGKSFREQ 632

RESULT 6
US-09-130-491-13
; Sequence 13, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491

22 EREVVVPIRLDP-----DINGRRYYW-----RGPEDSGDGLIFQITAFQ 61
49 EYEIVSPVRNALGEPPTNVHFKRTRRSINATDPWPAFASSSSSTSSQAHYRLSAFG 108
62 EDFYLHTPPDAQFLAPAFSTEHLGVP-----LOGLTGSSDLRRCFYSGDYNAPDSFAAV 117
109 QOFLNLTANAGFIAPLFTVTLTGGTGVNQTKFYSEAEALKHCFYK----- 155
118 SLGGGLRGAFYRGAEYVISPLPNASAPAAOQNSQGAHLQLLQRRGVPGGPDPTRSGVA 177
156 LRCQYQLRAHG-----RHQPLLANEHNKRSKDKKTRARKWGERINLAGDV 202
178 SGWNPAT--LRALDPYKPRRAGGESRRSRRAKRFVSPRYVETLVVADESVMKTHGA 236
203 AALNSGLATEAFSAVGKNTDNTREKTRHRT---KRELSYPRFEVLVAVADRNMVSYHGE 259
237 DLHYLLTLLTAAARLYRHPISILNPINTVVKVLLLRDSDGPKVTGNAALTLRNFCAWQ 296
260 NLQHYLTLSIVASIYKDPISIGNLINIVNLIVNHQDGPISISFNAQTLLKFNFCOWQ 319
297 KKLNVSKDHPEYWDTAIFLTRQDLGCA--TTCDTLGADVGTMCDPKRSCSVIEDDGLPS 355
320 HSNPGGIHH---DTAVLLTRQDICRAHDKCDTLGLAELGTICDPYRSCSISDSGLST 375
356 APTTAHELGHVFNPHDNVVKVEEFGKLRANHMSPILQIDRANPKSACSAALITDEL 415
376 APTTAHELGHVFNPHDDNNKKEE-GVYKSPQHVMAPTLNFYTNPMWMSKRSKYITEFL 434
416 DSGHGDCLLDOP--SKPISLPEDLPKASVYTLSSQCELAFAFGVSKPCPYMVOYCTKLWC---T 471
435 DTGYCECLLNEPESRYPPLVQLPCLLYNKNQXELIFGPGSQVCYPMQOCRLWCNNVN 494
472 GKAKQVCMQOTRHFHPADGTCGEGKCLKLGACVERHNLKHRVDGSKWAKWDPYGPISRT 531
495 GVHKG---CRTQHTPWADGTECEPGKCKXGFCVPK-EMDVPVTDGSGWSWSPGTCSRT 550
532 CGGGVQLARRQCTNTPANGKYGCVKRVKRSCLNLEPCPSSASGKSFPREQ 583
551 CGGGKTAIRECNRPKNKGGKVCVGRMRKFKSCNTEFLCKQK--RDFRDEQ 600

RESULT 8
US-09-369-364A-15
; Sequence 15, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 874
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-9
US-09-369-364A-15

Query Match 34.4%; Score 1087.5; DB 4; Length 874;
Best Local Similarity 43.1%; Pred. No. 2.1e-88;
Matches 217; Conservative 72; Mismatches 166; Indels 49; Gaps 12;

QY 115 AAVSLCGGLRGAFYRGAEYVISPLPNASAPAAOQNSQGAHLQLLQRRGVPGGPDPTRSG 174
Db 3 AVISLCGWMGTFRSHDGDYFIEPLQSVDEQEEDEEQNKPHIIRHSTPQREPSPTGKIAC 62
QY 175 GVASGNWP-----AILRA-----LDPYKPRRAGGESRRSR 205
Db 63 ATSELKNSHDKRRKTRMRKRRKRNLSLADVALLKSLATKVLISGYSNOT---NNTRDRW 119

269 VLLLRDRSDGPKVTGNAALTLRNFCAWQKLNKVSQKPEYWDTAIFLTRQDLGCAATCD 328
4 ILVHDEQKGEPTVSNAAALTLRNFCAWQKLNKVSQKPEYWDTAIFLTRQDLGCAATCD 63
329 TLGMADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVVKVEEFGKLRANH 388
64 TLGMADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVVKVEEFGKLRANH 123
389 MMSPTLLQIDRANPWSACSAALITDLSGHDGCLLDOPSKPISLPEDLPKASVYTLSSQOC 448
124 MMASMLNLDHSPQWSPCSAYMITSFLDNGHGECLMDKPNPQLPGLDPTSYDANROC 183
449 ELAFVGSKPCP--YMQVCTKLWCTGAKQMGVQOTRHFHPADGTCGEGKCLKLGACVER 507
184 QFTFGEDSKHCPDAASTCSLWCTGSGVILVQCTRHFHPADGTCGEGKCLKLGACVER 243
508 HNLNKH---RVDGSAKWDYGPSCRTCGGGVQLARRQCTNTPANGKYGCVKRVKRS 564
244 TD-RKHFDTPFHSGWMGWPBGDCSRTCGGGVQYTHRECDNPVKNKGGKCEGKRVYRS 302
565 CNLEPCPSSASGKSFPREQ 583
303 CNLEDCPDN-NGKTFREQ 320

RESULT 7
US-09-369-364A-13
; Sequence 13, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1882
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-9
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (468)
; OTHER INFORMATION: Xaa = C
; NAME/KEY: MOD_RES
; LOCATION: (521)
; OTHER INFORMATION: Xaa = Y
US-09-369-364A-13

Query Match 35.4%; Score 1121; DB 4; Length 1882;
Best Local Similarity 40.0%; Pred. No. 6.2e-91;
Matches 237; Conservative 86; Mismatches 199; Indels 70; Gaps 15;

[illegible]

```

RESULT 9
US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirschata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7

```

```

Query Match      29.5%; Score 932.5; DB 4; Length 997;
Best Local Similarity 39.3%; Pred. No. 1.7e-74;
Matches 249; 'Conservative' 69; Mismatches 217; Indels 99; Gaps 29;

QY      1  MILLIGTILLAFAGRTAGGFEPER---EWVVPRLDP-----DINGRR- 39
      :|||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      18  LULLICALPAGPAPGAPGRATEGRAALDIVHPVRVDAGGSFLSYELWPRALRKRDVSRRD 77
      :|||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      40  ---YY---WRPESGDGLIFQTAFQDFYHLHLPDAQFLAPAF--STEHLGVPLQGL 91
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      78  APAFELOYRGEE-----LRFNTAQ-----HL-----LAPGVSTRRRG-----GL 116
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      92  TCGSSDLR-----RCFYSGDVNAEPD---SFAAVSLCGLRGAFGYRGAETVPSLPNASHA 144
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      117  --GRAHIRAHTPACHLLGEVQ--DPELGGGLAAISACDGLKGVQLSNEDYFIEPLD--SA 171
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      145  PAAQRNSOGAHLLQRGVV-----GGPSGDPTSRGCVASGWNPAILRALDLPYKPRRAGFG 199
      || :| :| | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      172  PARPQHQAQ--PHYVYKQAPERLAQRGDSSAP--STCGV-----QVYPELESRRERW 219
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY      200  ESRSRRSRGRKRF---VSIPTRYVELTVADESNWKFHG--ADLEHYLLITLTLATARIYR 254
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      220  EQQQWRAPRLRLRHQRSVSKKCEFTLVADAKMVEYHGHPQGVESYIVTINMWAGFLH 279
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

QY	255	HPSLNPINIVVVKVILLRRDRSGPKVTGNAALTLRNFCAMOKKLNKVS DKHPEYWDTAI	314
Db	280	DPSIGNPIHITIIVRLVLEDEEDLKITHUADNTLKSFKWKWSINKMGDAHPLHHDTAI	339
QY	315	LFTRODLGGATT--CDTLGMADVGTWCDPKRSCSVTEDDGLPSAFTTAHELGHVFNMPHD	372
Db	340	LLTRKDLCAAANRPCTELGSHVAGMCQPHRSCSINEDTGLPFAFVAHELGHSEFIQH	399
QY	373	-NVKCEEVFGKLRAHNMSPTLIQIDRANPWASACSAAITFDLODSHGDCLLDQPSKP-	430
Db	400	GSSNDCEPV-GK--RPFIMSPQLLYDAAPLTWSCRQRYITRFLDRGWGCLDDPPAKDI	456
QY	431	ISLPEDLPAGASYTLSOOCELAFGVGSKPCPYM-QYTCKLWCTGKAKGMVCOTRHFPWAD	489
Db	457	IDFPSVPVGLYDVSHQCRLOYGAYSAFCEDMDNVCHLMCSVGT-----TCISKLDAAVD	512
QY	490	GTSCEBGKCLKGACVERHNLRHYDVGGSWAKWDPYGPCSRTOCGGVQLARQCNTPTPA	549
Db	513	GTRCENKNWCLSGECPV-VGFEPFEAVDGGWGSGWSANSICRSOGMGVQAERCQTCTPTK	571
QY	550	NGKKYCEGVKVYRSCNLEPCPPSSASGKSFREHQ	583
Db	572	YKGRCYCVGERKREFLCNLQACP--AGRPSFRHVQ	603

RESULT 10
 US-09-491-522-5
 ; Sequence 5, Application US/09491522
 ; Patent No. 6428998
 ; GENERAL INFORMATION:
 ; APPLICANT: Colige, Alain
 ; APPLICANT: Lapierre, Charles M.
 ; APPLICANT: Prockop, Darwin J.
 ; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
 ; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds, LLP
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10036-2811
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows
 ; SOFTWARE: Fastseq for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/491.522

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886,333
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8389-0060-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-491-522-5

Query Match	24.7%;	Score 782.5;	DB 4;	Length 1211;
Best Local Similarity	34.9%;	Pred. No. 5.2e-61;		

Matches 217; Conservative 78; Mismatches 247; Indels 79; Gaps 27;

QY 9 LAFAGTAG--GFEPEREVVPRLDPD-----INGRR-----YYW 42
Db 39 LAAADPPGPGPGHGAERILAVPVTDAQRLVSHVSAANTSRAGVARRAPVTSFP 98
QY 43 RGPEDSGOGLFIQTAFQEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGSSDLRCF 102
Db 99 GGNEEPEGSHLYNVTVFGRDLHLRPNARLVAPGATMEWQGE--KGTTRVEPLGSL 156
QY 103 YSGDYN--AEPDSFAVSLCGLRGAFYGAAYVISPPLNASAPAAQNSQG-AHLQR 159
Db 157 YVGDVAGLAESS-VALSNCGLAGLIRMEEEFIELEKGL--AAQEAEGQVHVYR 213
QY 160 RGVPGGSDPTSRGCGVAGWNPATLRALDPYKPRRAGFSGESRRSRRAKRFVSPRY 219
Db 214 RPPTSPPLGPPQA---LDTG---ASLSDLSL-RALGVLEEHANSRRRARRHAADDY 266
QY 220 VE-TLVVADESVMKPHGAD-LEHYLLTLTAARLYRHPSTILNINIVVVKLLRRDS 277
Db 267 NEIVLLGVDDSVVQFQKHEVQKYLTLMLNIVNEYIHDESGLAHINVVLRILLSYGS 326
QY 278 GPKV-TGNAALTRNFCAWOKLNKVSQKHPEYWDATILFTRODLCGATTCDTLGMADV 336
Db 327 MSLEIENPQSLENCWAYLQRPDPDGHDEYHDAFLTRQDP-GPSGQ--GYAPVT 383
QY 337 TMCDPKRSVIEDDGLPSAFTTAHELGHVFNMPHDNV-KVCEEVFGKLRANHHMSPTLI 395
Db 384 GMCHPVRSTLNHEDGSSAFVAHETGVLGMEHDGQNRCD--EVRLSIMAPLVQ 440
QY 396 QIDRANPWSACAAIITDPLSGHGDCLLDQ---SKPISLPEDLPASVYLSQOCELA 452
Db 441 AAFHRFHSRCSQBELSYLHS--YDCLLDQPFADWDP-ALPQ-LPGLHYSMNEQCRDF 496
QY 453 GVSRSKPCPYMQY---CTKLWCTGKAKGMVQTRHFWADTSCGEGKCLKGACV---- 505
Db 497 GLGYMCTAFRTFDPCKQLWCS-HPDNPYFCKTKGPDLDGTMCAPGRKCFKHCINWTP 555
QY 506 ---ERHNLKRVDSWAKWDPYGPCSRTCGGQVQALARRQCTNPTTPANGRYKCVRVKY 562
Db 556 DILKR-----DGSWGAWSFGSCSRTCGTGVKFRTRQDNPHPANGRTCSGLAYDF 607
QY 563 RSCNLEPCPSSASGSFREQ 583
Db 608 QLCRSRQDCPDSLA--DFREEQ 626

RESULT 11
US-09-930-872-4
; Sequence 4, Application US/09930872
; Patent No. 6448388
; GENERAL INFORMATION:
; APPLICANT: Fridge, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. 6448388a1 Human Proteases and Polynucleotides Encoding the Sa
; FILE REFERENCE: LEX-0219-USA
; CURRENT APPLICATION NUMBER: US/09/930.872
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,852
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1224
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-930-872-4

Query Match 24.1%; Score 763; DB 4; Length 1224;
Best Local Similarity 34.0%; Pred. No. 2.9e-59;
Matches 200; Conservative 74; Mismatches 227; Indels 88; Gaps 21;

QY 51 QGLIFQITAFQEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGSS-----DLRRCFY 103

Db 95 ESLHLRLKGRHDFHVDLRTSSSLVAPGFIVQTL-----GKTGKSVQTLPPDEF--CFY 147
QY 104 SGDVNAEPDSFAVSLCGLRGAFYGAAYVISPPLNASAPAAQNSQG---AHLQRR 160
Db 148 QGSLRSHRNSVALSTCQGLSGMIRTEADYFLRPLPSHLWKLGRAAQSPPSHVLYKR 207
QY 161 G----VPGGSGDPTSR-----CGVASGWNPA-----ILRA 187
Db 208 STEPHAPGASEVLVTSRTWELAHQPLHSSDLRLGLPQKHQFCGRRKKYMPQPKEDFL 267
QY 188 LDYKPPRAGFSGSRSSRRSRRAKRFVSPRYVETLVVADESVMKPHG-ADLEHYLLTL 246
Db 268 PDEYK---SCLRHRKSLRSHRNEEL-----NYETLVVVDKMMQNHENITTYVLTIL 319
QY 247 ATARLYRHPSTILNINIVVVKLLRRDSGPKVTGNAALTLRNFCAWOKKL-NKVSOK 305
Db 320 NWSALFKQGTIGTNINIAIVGLILLEDEQPLVISHHADHTLSSFCQWQSGMLGKDGTR 379
QY 306 HPEYWDATILFTRODLG---ATTCDTLGMADVMTCDPKRSCSVIEDDGLPSAFTTAHEL 363
Db 380 H----DHAILTLGLDICSWNKNEPCDTLGFAPISGMCYSKRSCTINEDTGLGLFTIAHES 435
QY 364 GHVFNMPHDNV-KVCEEVFGKLRANHHMSPTLIQIDRANPWSACSAIITDPLDLSHGDC 422
Db 436 GHNFMTIHDGEGNMCKSEG-----NIMSPTLAGRNGVFSWSPCSROYLHKFLSTAQAIC 490
QY 423 LLDQPSKPI---SLPEDLPASVYLSQOCELAFGVGSKPCPY---MQYCTKLWC---TGKA 474
Db 491 LAQDP-KPVKEYKYPEKLPGLDYDANTQCKWQGEKAKCLMDLDFKDKICKALWCHRIGRK 549
QY 475 KGQMVQCTRHPFWADTSCGEGKCLKGACVERHNLKRVDSWAKWDPYGPCSRTCGG 534
Db 550 ----CETKMPAAEGTICGHDMMWCGQCVKYGDGPKPTHGHWDSWSSWSPCSRTCGG 604
QY 535 GVQLARRQCTNPTTPANGRYKCVRVKYRSCNLEPCPSSASGSFREQ 583
Db 605 GVSHRSRLCTNPXKFGKCEGCEGSTRTLKLCNSQKCPDPS--VDFRAAQ 651

RESULT 12
US-09-369-364A-17
; Sequence 17, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-10
US-09-369-364A-17

Query Match 23.7%; Score 750.5; DB 4; Length 1081;
Best Local Similarity 35.0%; Pred. No. 3.2e-58;
Matches 191; Conservative 73; Mismatches 235; Indels 47; Gaps 19;

QY 64 FYLHLTPDAQFLAPAFSTHGLVPLQGLTGSSDLRRCFYSGDVNAE-PDSFAVSLCGG 122
Db 65 FLNLTRSSRLLAGRVSYEW--TREGLAWQRAARPHCLYAGHLQQAQSSHVAISTCGG 122
QY 123 LRGAFGYRGAEYVISPPLNASAPAAQNSQAHLLORRGVPGGSDPTSRGCV--ASGW 180
Db 123 LHGLIVADEEYLLIEPLHGGPKGSRSPESGPHCVKRSRLRHPHLD--TACGVREKWP 180
QY 181 --NPAILRALDPYKPRRAGFSGESRRSRRAKRFVSPRYVETLVVVADESVMKPHG-AD 237

Db 181 KGRPWRLTKP--PPARPLGNTERGPG-LKRSVSREYVETMDVADKMMVAYHGRD 237
QY 238 LEHYLTLLATARLYRHSILNPNIVVVKVLLLRDRSGPKVTGNAALTLRNFCAWQK 297
Db 238 VEQYVLAIVNIVAKLFODSSLGSTWNILVTRILLTEOPLTEIHHAGKSLDSFCKWQK 297
QY 298 KL-----NKVSDRHPYWDTAILFTRODLC--GATTCDTLGMADVGTMDCKPRSCSVI 348
Db 298 SIVNHSCHGNAIPENGVAHNDTAVLITRYDICIYKNKPCGTLGLAR-WAECVSAREAAAS 356
QY 349 EDGLSATSATTAHELGHVFNPHDNVYCEEVFG-----KLRAHNMSTLTIQIDRNP-- 402
Db 357 MRLTAATSVHHCEIGHTFGMHNDGVNCGARGQDPAKLAHAHITMKT-----NPFV 409
QY 403 WSACSAIITDFDSHGDCLLDOPSK-PISLPEDLPAGASTYLSQOCELAFVGSVKPCPY 461
Db 410 WSSCNRDYITSLFDSGLGLCLNRPQDFVYPTVAPGQAYDADQCRFQHGKSRCKY 469
QY 462 MOYCTKLWCTGRKAGQWVQTRHFPWADTSCGEGKL----CLKGACVERHNLKHRVDG 517
Db 470 GEVCESELMCLSKSNR---CITNSIPAAGTLCQTHIDKGMCKYKVCVP-FGSRPEGVDG 525
QY 518 SWAKWDPYPCSTRCTGGVQLARRQCTNPTPANGKYCEGVKVKYRSCNLEPCPSSASGK 577
Db 526 AWGPWTPWGDSCSTRCTGGVYSSSRHCDSPRPTIGGKYCLIGERRHRSNTDDCPPGS--Q 583
QY 578 SFREQ 583
Db 584 DFEVQ 589

RESULT 13

US-09-491-522-11
; Sequence 11, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B.
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1205 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-11
Query Match 23.4%; Score 739; DB 4; Length 1205;
Best Local Similarity 32.5%; Pred. No. 3.9e-57;
Matches 212; Conservative 81; Mismatches 242; Indels 118; Gaps 28;
QY 1 MLILGLITLAFAGRTAG-----GFEPEREVVPPIRLD----- 32
Db 16 LLLLLLLPLPADARLAAAAADPPGPGQGHGAERILAVPRTDAQGRVSHVVSAAAPAG 75
QY 33 -----PDINGRRYWRGPDSDGQGLIFQITAFQEDFYHLTPDQAFAPAFST 81
Db 76 VTRRAAPAQIFCLSG-----GSEEDPGRLFYNTVFGRLHLRLFRNARLVAPGATV 129
QY 82 EHLGVPLQGLTGSSDLRRCFYSGDV-NAEPDSFAAVALCGGLRGAFGYRGAEVVISPLP 140
Db 130 EWQGE--SGATRVPEPLLTCLYGVGVAGLAESSVALSNCGLAGLIRMEEEFFIELE 187
QY 141 NASAPAAQNSQG-AHLORRGVPGGPGSDPTSR-----CGVASGNWPAIRAL 188
Db 188 KGL--AAKEAEQGRVHVYHFT-----TSREFFLGGFQALDTGISADSLSLRAL 237
QY 189 DPYKPRRAGFESRRSRGRKRFVSIPIRY-VETLVVADESVMKFGHAD-LEHYLTLL 246
Db 238 -----GVLEERVNSRRRRHAADDDYNIIEVLGVDDSVVQFHGTEHVQKYLITLM 289
QY 247 ATAARLYRHSILNPNIVVVKVLLLRDRSGPKV-TGNAALTLRNFCAWKKLNKVSXK 305
Db 290 NIVNEYIHDESIGAHINVLVRIILLSYGKMSLIEIGNPSQSLNVNCRWAYLOQKPTD 349
QY 306 HPEYWTAILFTRODLCGATTCDTGLMADVGTMDCKPSCSVIEDDGLPSAFTTAHELGH 365
Db 350 HDEYHDAIFLTRQDF-GPSGMO--GYAPVTGCHPVRSCITLNHEDGFSFAVVAHETGH 406
QY 366 VFNNPHDNV-KVCEEYFGKLRANHMSPPTLIQIDRANPWSACSAIITDFDSHGDCLL 424
Db 407 VLGMEDHGGQNRCD--EVRLLGSIMAPLVQAARHFRHWSRCSQOELSRYLHS--YDCLR 461
QY 425 DQP--SKPISLPEDLPGASYTLSSQOCELAFVGSVKPCPYMAY---CTKLCTGKAKGQM 478
Db 462 DDPFTHDWP-ALPQ-LPGLHYSMNEQCRDFGLGTYMMCTAFTFPDCKQLWCS-HPDNPY 518
QY 479 VCQTRH-FPWADTSCGEGKLCLKGAC-----VERHNLKHKRVYDGSWAKWDPYGPCSR 530
Db 519 FCKTKKGFP-LDGTMCAPGKHCFCGHCTWLTDFILKR-----DGNKGAWSPFGCSR 569
QY 531 TCGGGVQLARRQCTNPTPANGKYCEGVKVKYRSCNLEPCPSSASGKSFREQ 583
Db 570 TCGTVKFRTRQCDNPHANGGRTCSGLAYDFOLCNSQDCPDALA--DFREQ 620

RESULT 14

US-09-369-364A-5
; Sequence 5, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Aptel, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-6
; FEATURE:

```
; NAME/KEY: MOD.RES
; LOCATION: (450)
; OTHER INFORMATION: xaa = L
; US-09-369-364A-5

Query Match      22.5%; Score 712.5; DB 4; Length 859;
Best Local Similarity 30.2%; Pred. No. 5.7e-55;
Matches 183; Conservative 89; Mismatches 215; Indels 119; Gaps 18;

Qy 24 EVVPIRLDP-----DINGRRYWRGPED--SGDQGLIFQITAFQEDFYHLHP 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 43 QLTPIRVDQNGAPLSEFTVKNDRKHSRRSMDPIDPQAVSKLFFKLAYSAGKHFLNLTL 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 71 DAQFLAPAFSTEHLGVPLQGLTGSSDLRFCEYSGDVNAEPDSF-AAVSLCGGLRGAFGY 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 103 NTFVSKHFTVEYWG--KDGQWQKHDLDNCHTYGLQDORSTTKVALSCVGLHGVIAI 160
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 130 RGAEYVISPPLNASAPAAQRNSQAH-----LLQRRGVPGGSDPTSRGCVASGWN 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 EDEYFIEPLKNTTEDSKHFSYENGHPHYVKKSLAQORHL-----YDHSCHGVSD--- 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 182 PAIRALDOPY-----KPRAGGESRRSRRRAKRFVSIPIRYVETLVVADESVMKFG-A 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 --FTRSGKPMWLNDTPLFLIHYQINNTIHHROKRSVSIERFVETLVVADKMMVYHGRK 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 237 DLEHVLLTLATARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCWQ 296
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 DIEHYILSVMNIVAKLYRDSLSLGNVNLIVARLIVLFEDQPNLEINHADKSLDSFCKWQ 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 297 KKL-----NKVSKHPEYWDTAIFLTRODLGATT--CDTGMADVGTMDPKRSCSV 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 330 KSILSHQSDGNTIPENGIAHDNAVLTRYDICTYKKNKPGTGLASVAGMCEPERSCSI 389
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 348 IEDGLPSAFTAHELGHVFNHMDNVKVEEFGKLRANHMSPTLIQIDRANPWSACS 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 390 NEDIGLSAFTIAHEIVHNFNMHNDGI-----GNSCG 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 408 AAITFDLSDGHGCLLDQPSKPSLDPGLPGASYITLSQOCE-----LAFGVGSKPCPY 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 422 RYVKQ-----QNYGSSHCEYQSFYLVCLQSRXHHQLF 455
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 462 MOYCTKLCTKAKQMWVQOTRHFPMADGTSCGEGKL-----CLKGACVERHNLNKHVDG 517
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 456 REVCRELWCLSKSNR---CVTNISIPAAEGTLQCTGNIEKGWCYQDGVF-FGTWPQSIDG 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 518 SWAKWDPYPCSCGCGGVQLARQCTNPTPANGKYCEGVVRYRSCNLEPCPSSASGK 577
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 GWGPNLWGECSRTCGGCVSSSLRHCDSPADSEVEKYCLGERKRYRSCNTDPCP--LGSR 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 578 SFREQ 583
   : : : : :
Db 570 DFEKQ 575
   : : : : :

RESULT 15
US-09-130-491-16
; Sequence 16, Application US/09130491
; Patent No. 6416574
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 551
```

```
; TYPE: PRT
; ORGANISM: Rattus rattus
; US-09-130-491-16

Query Match      21.7%; Score 687; DB 4; Length 551;
Best Local Similarity 56.4%; Pred. No. 5.8e-53;
Matches 119; Conservative 29; Mismatches 57; Indels 6; Gaps 4;

Qy 377 CEEVFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDSGHGDCLLDQPSKPSLSPED 436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 CASLNGVSGDSHLMASMLSSLDHSQSPWSPCSAYMYVTSFLDNHGCECLMDKPKQNPILPSPD 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 437 LPGASYTILSQOCELAFGVGSKPCP-IMOYCTKLWCTGKAKGQMVQOTRHFPMADGTSCGE 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 LPGTLYDANRQCOFTFGEESKHCPDAASTCTTLWCTGTSGLLVCQTGKHFPMADGTSCGE 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 496 GKILCLGACVERHNLNKH---RVDGSAKWADPYGPCSRTCGGVQLARQCTNPTPANGG 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 GKWCYSGKCVNKTDM-KHFATPVHSGWGFPGWGDGCSRTCGGVQYTWRECDNPVPKNGG 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 553 KYCEGVVRYRSCNLEPCPSSASGKSFRREQ 583
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 KYCEGKRVYRSCNIEDCPDN-NGKTFREQ 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: April 29, 2003, 17:11:51
Job time : 13.4569 secs
```